DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA 91125

THREE PROBLEMS OF THE THEORY OF CHOICE ON RANDOM SETS

B. A. Berezovskiy,* Yu. M. Baryshnikov, A. V. Gnedin

Institute of Control Sciences
Academy of Sciences of the U.S.S.R., Moscow

*Guest of the California Institute of Technology

SOCIAL SCIENCE WORKING PAPER 661

December 1987

THREE PROBLEMS OF THE THEORY OF CHOICE ON RANDOM SETS

B. A. Berezovskiy, Yu. M. Baryshnikov, A. V. Gnedin

Institute of Control Sciences
Academy of Sciences of the U.S.S.R., Moscow

ABSTRACT

This paper discusses three problems which are united not only by the common topic of research stated in the title, but also by a somewhat surprising interlacing of the methods and techniques used.

In the first problem, an attempt is made to resolve a very unpleasant metaproblem arising in general choice theory: why the conditions of rationality are not really necessary or, in other words, why in every-day life we are quite satisfied with choice methods which are far from being ideal. The answer, substantiated by a number of results, is as follows: situations in which the choice function "misbehaves" are very seldom met in large presentations.

In the second problem, an overview of our studies is given on the problem of statistical properties of choice. One of the most astonishing phenomenon found when we deviate from scalar-extremal choice functions is in stable multiplicity of choice. If our presentation is random, then a random number of alternatives is chosen in it. But how many? The answer isn't trival, and may be sought in many different directions. As we shall see below, usually a bottleneck case was considered in seeking the answer. It is interesting to note that statistical information effects the properties of the problem very much.

The third problem is devoted to a model of a real life choice process. This process is typically spread in time, and we gradually (up to the time of making a final decision) accumulate experience, but once a decision is made we are not free to reject it. In the classical statement (i.e. when "optimality" is measured by some number) this model is referred to as a "secretary problem", and a great deal of literature is devoted to its study. We consider the case when the notions of optimality are most general. As will be seen below, the best strategy is practically determined by only the statistical properties of the corresponding choice function rather than its specific form.

THREE PROBLEMS OF THE THEORY OF CHOICE ON RANDOM SETS

B. A. Berezovskiy, Yu. M. Baryshnikov, A. V. Gnedin

Institute of Control Sciences
Academy of Sciences of the U.S.S.R., Moscow

Introduction.

The three problems discussed below are united not only by the common topic of research stated in the title, but also by a somewhat surprising interlacing of the methods and techniques used. And this interlacing is indeed surprising, since the problems treat quite different things.

In the first problem, an attempt is made to resolve a very unpleasant metaproblem arising in general choice theory: why the conditions of rationality are not really necessary or, in other words, why in our every-day life we are quite satisfied with choice methods which are far from being ideal. The answer, substantiated by a number of results, is as follows: situations in which the choice function "misbehaves" are very seldom met in large presentations.

In the second problem, an overview of our studies is given on the problem of statistical properties of choice. One of the most astonishing phenomenon found when we deviate from scalar-extremal choice functions is in stable multiplicity of choice. If our presentation is random, then a random number of alternatives is chosen in it. But how many? The answer isn't trival, and may be sought in many different directions. As we shall see below, usually a bottleneck case was considered in seeking the answer. It is interesting to note that statistical information effects the properties of the problem very much.

The third problem is devoted to a model of a real life choice process. This process is typically spread in time, and we gradually (up to the time of making a final decision) accumulate experience, but once a decision is made we are not free to reject it. In the classical statement (i.e. when "optimality" is measured by some number) this model is referred to as a "secretary problem", and a great deal of literature is devoted to its study. We consider the case when the notions of optimality are most general. As will be seen below, the best strategy is practically determined by only the statistical properties of the corresponding choice function rather than its specific form.

The three problems feature, in our opinion, some interesting methodological peculiarities. The first problem makes use of some more realistic description of the choice situation, i.e. of a more adequate model to overcome a certain contradiction in the intuitive perception of the choice situation.

The problem treating the optimal statement shows how the study of particular cases allows derivation of general regularities, while the problem on statistical properties of the choice function gives evidence of how some particular constrained situations may distort the real life picture.

Thus, all three problems are actually three fragments of the theory of choice on random sets being developed by the authors. As is the case with all newly-born theories, the number of problems unsolved increases with our knowledge.

ASYMPTOTICAL EQUIVALENCE OF CHOICE FUNCTIONS

Introduction. Classical Choice Model (*Choice on Deterministic Sets*).

The choice theory is a theory of our preferences. Our preferences, roughly speaking, imply an ability to choose the best from a set of different variants (alternatives).

This idea underlies the notion of a choice function. This is a composite notion. To specify a choice function one should first of all outline a set of objects which can be presented for a choice at all. In what follows, such objects will be denoted as U (Fig. 1.1, a).

Second, we have to outline sets of objects presented from which we can choose. These sets will be referred to as admissible presentations, or, simply, presentations. We denote them as $\chi \subset 2^U$. And, finally, the main thing we have to do is to define the choice function C, which relates to each subset $X \in \chi$ its chosen part $C(X) \subseteq X$.

Such triple (U, χ, C) is the central object of the classical choice theory. Let us give some examples of well known choice functions (Fig. 1.2).

1. Classical Scale-Extremal Choice Function

   Let $f: U \to \mathbf{R}$ map the set of alternatives U into real line $\mathbf{R}$, and the optimality principle is specified by the condition: x is better than y if $f(x) > f(y)$ and the variant to be chosen should feature the maximum value of f among the alternatives presented, so that

   $$C^E(X) = \{x \in X: \ \forall y \in X \ f(x) > f(y)\}.$$

   This is an example of the most widely used function, treated in great detail in mathematical programming.

2. Assume that $U \subset R^n$. This is true in the Pareto choice function

   $$C^{Par}(X) = \{x \in X : \forall y \in X, y \neq x \ \exists i: \ y_i < x_i\}$$

   i.e. point x is chosen if any other point $y \in X$ has at least one coordinate value less than the corresponding coordinate value of x.

3. Plott's choice function, also referred to as the Collected-Extremal choice function ($U \subset R^n$):

   $$C^{Pl}(X) = \{x \in X: \ \exists i: \ \forall y \in X \ x_i > y_i\}$$

   i.e. a point is chosen if it has a maximal value in at least one of its coordinates.

4. Yet another widely used choice function - Tournament choice function. A formal definition of the tournament choice function is as follows: let $f: U \times U \to \mathbf{R}$ be defined on pairs of variants (x_i, x_j). It is implied that $f(x_i, x_j)$ are the scores gained by x_i in the games against x_j. For every element $x_i \in X$ consider the sum

   $$m_X(x_i) = \sum_{l=1}^{k} f(x_i, x_l)$$

which defines the score gained by x_i in tournament X. Then the tournament choice function prescribes that the variant(s) with maximum score m_X should be chosen.

Below we shall also consider a modification α of the Tournament choice function. α-Tournament choice function is one which prescribes the selection of α-share of the variants from X with the highest tournament score.

These choice functions are well known and widely applicable in numerous optimization models. We regard this fact as highly remarkable because wide applicability of the above choice functions indicates that we intuitively accept them. At the same time, we accept some choice principles which are known as rational ones since they make sense to us.

Let us recall two of these principles (Fig. 1.3). We say that a choice function satisfies the heritage (H) condition if a variant chosen in some set is chosen in any of its subsets too:

$$\forall X' \subset X \in \chi,\ C(X') \supseteq C(X) \cap X'.$$

Another principle is expressed by concordance (C) condition:

$$\forall X, Y \in \chi,\ C(X) \cap C(Y) \subseteq C(X \cup Y)$$

i.e. if a variant x is chosen both in X and Y, it is chosen in $X \cup Y$ as well.

It turns out that not all of the choice functions satisfy intuitively reasonable choice principles.

Indeed, it is easy to see that in the first and the second examples, the choice functions satsify **H**- and **C**-conditions. The Collective-Extremal choice function does satisfy **H**- condition, but does not satisfy **C**-condition. The Tournament and α-Tournament choice functions fail to satisfy any of these conditions. This paradox and its resolution is the main subject of the present problem.

Extended Choice Model (*Choice on Random Sets*)

Let us go back to the description of a choice situation to show that our choice model was incomplete (Fig. 1.1 a, b).

Indeed, each choice function is characterized by the presence of a decision maker and another party, or factor, presenting or forming the choice alternatives. If we ignore this other party or factor, then in the analysis of the choice function we will have to follow its behavior in all admissible presentations. Obviously the general choice function is complex, and the theory inevitably falls into the framework of the combinatorial ideas going more and more apart from the real world of human preferences. In order to facilitate the interpretation of the results obtained, and bring them closer to the intuitive ideas on optimality that a human being posesses, it is necessary to take into account structure of the presentations from which the choice is made.

We shall try to demonstrate this by introducing a probabilistic structure on a set of presentations, in other words, the subject of our investigation is the choice functions on random presentations.

Two more essential features are inherent in the theory which follows. First: we study the taxonomy (systematics) rather than "physiology" of choice functions, i.e. we are not trying to find out which choice functions are constructed of which and what mechanisms determine these functions. We want to establish the similarity between choice functions which may show a different structure but, nevertheless, be practically the same from a consumer's viewpoint. In other words, we wish to determine a certain proximity been choice functions, and will concentrate upon this problem.

Second: considered will be feasible finite presentations of a large size.

Pair-Dominant Choice Functions

Very often our interpretation of optimality is related with the concept of preference, i.e. with the ability to make pairwise comparison of the alternatives to see which is the best, if any. Mathematically, the preference is specified by binary relation R (x is better than y if $(x,y) \in R$) (Fig. 1.3). Each binary preference relation is associated with the so-called pair-dominant choice function:

$$C_R(X) = \{x : \nexists\, y : yRx\}$$

Pair-dominant choice functions associated with various binary preference relations are very popular. For instance, the problem of maximizing a scalar function f is reduced to such a choice function: here

$$xRy \Leftrightarrow f(x) > f(y).$$

Another example of a pair-dominant choice function is Threshold choice function C^B: in a set U a subset B is isolated and exactly those alternatives in X are choosen which are found in the subset B. It is quite evident that C^B is associated with C_R in which

$$R = \{(x,y) \in U \times (U/B)\}.$$

Such a choice function appears, for example, in quality control (on a Go-NoGo basis).

All pair-dominant choice functions satisfy the **H**-property (Fig. 1.4). Taking a look in reverse direction, we can observe a following remarkable fact: any choice function with the **H**-property can be presented in the form of a union of, perhaps, an infinite number of pair-dominant functions.

Below we shall mostly deal with the case when the choice function featuring the **H**-property is decomposable into a finite number of pair-dominant choice functions.

Indeed, each pair-dominant choice function C_{Ri} reflects some preference "hidden" in the decision maker's concept of optimality. Clearly, even such a perfect creation as a human mind cannot hold anything but a finite number of optimization criteria. Let us consider the property which guarantees this finiteness.

Assume that the choice function features property $\mathbf{C}_k$ if for any k admissible presentations $X_1, \ldots, X_k$, a point chosen in each X_i $(i = 1, \ldots, k)$ is chosen in the union of any two of them. For $k = 2$ this condition, $\mathbf{C}_2$, coincides with the classical condition of concordance, **C**.

Assertion. If the function features both the **H**-condition and property $\mathbf{C}_k$, then it can be presented in the form of a union of no more than $k - 1$ pair-dominant choice functions.

These have been all auxiliary data from the abstract choice theory that will be useful below.

Now we come down to a strictly formal mathematical model of choice from a random presentation.

Basic Objects (*Fig 1.5*)

So, U is a total set of alternatives, and C is a choice function on U. We assume that any finite set is admissible. To compare the choice results, we introduce a metric on a set of finite subsets U:

$$\partial(X,Y) = \frac{\mathrm{card}(X \Delta Y)}{\mathrm{card}(X \cup Y)}$$

We omit proving* that ∂ is indeed a metric. The calculations are cumbersome but trivial. We only note the following property: $0 \le \partial \le 1$ and $\partial = 1 \Leftrightarrow X \cap Y = \theta$.

The metric ∂ is a natural measure of difference between two sets. Intuitively this may be understood this way: if ∂ is small then choosing randomly an element of one set yields with a high probability that this is also an element of the second set and vice versa.

Like we already mentioned, we are dealing with finite random subsets of large cardinality, i.e. we are interested in the asymptotic behavior of the choice function in tending the number of variants to infinity.

Let (θ) be a family of finite random subsets where θ is a real parameter ($\theta \in \mathbf{R}$). This implies that for each θ a finite random subset is defined. We will assume that with $\theta \to \infty$ the cardinality of this subset (provided it is measurable) tends to infinity.

Definition. Two choice functions are called asymptotically equivalent with respect to a family $X(\theta)$, if $\partial(C_1(X(\theta)), C_2(X(\theta))) \xrightarrow{P} 0$ with $\theta \to \infty$.

The definition implies that two choice functions are asymptotically equivalent if the choices made from subsets large enough with probabilities as close to one as desired show an arbitrarily small difference.

Note that the definition of the asymptotic equivalence is essentially dependent upon family $X(\theta)$, a random subset. When the family changes, the asymptotically equivalent functions may become non-equivalent, and vice versa.

---

*However we omit all the proofs. You can find them in Baryshnikov, Y. M., Berezovskiy, B. A., Borzenko, V.I., Kempner, L. M., "Multicriterial Optimization: Theory and Applications," *Optimization Software*, New York (to appear in 1988) and Berezovskiy, B. A., Gnedin, A. V., "The Best Choice problem." *Nauka*, Moscow, 1984.

Basic Constructions (Fig. 1.6)

Consider some other notions which are also essentially dependent upon the family $X(\theta)$.

We say that functions C_1 and C_2 are asymptotically independent if the mean cardinality of intersection of the choices is asymptotically small as compared to the mean cardinality of each choice, as $\theta \to \infty$, i.e.

$$\frac{E\,\mathrm{card}(C_1(X(\theta) \cap C_2(X(\theta)))}{E\,\mathrm{card}(C_i(X(\theta))} \to 0, i = 1,2 \text{ with } \theta \to \infty$$

Choice function C will be called well-conditioned if the variance of the choice cardinality is asymptotically small as compared to the square of the mean cardinality of choice, or in other words, if the relative deviation of the choice cardinality from the mean is small enough as $\theta \to \infty$.

In our further considerations we will make use of the fact that the relation of the asymptotic equivalence on choice functions is indeed an equivalence relation, i.e. C_1 aeq C_2 and C_2 aeq C_3 implies that C_1 aeq C_3; C_1 aeq C_1 and C_1 aeq C_2 implies that C_2 aeq C_1. This may be easily derived from the fact that ∂ is a metric. Let us outline a class of random subset families dealt with below.

Let a probabilistic measure ν be specified on U such that the measure of each point equals zero (this assumption is made to avoid manipulation with multisets). Assume that $X(n)$ is a repeated independent sample of size n. Here n has the meaning of parameter θ.

Note that to have the pair-dominant choice cardinality measurable in such a class of random subsets, it is sufficient to have the binary relation $R \subseteq U \times U$ measurable, whereas for measurability of the cardinality of choice functions constructed from the pair-dominant ones it is sufficient that the respective binary relations are measurable.

Let us now formulate an auxiliary theorem serving as an important theoretical tool to establish the asymptotic equivalence.

Auxiliary Theorem. Let $C_1 \subseteq C_2$, with probability 1 let the ratio of mean cardinalities of choice over these two functions tends to unity, as n tends to infinity, and let one of these two choice functions be well-conditioned. Then these two choice functions are asymptotically equivalent.

Note that the requirement of being well-conditioned is quite essential here: although mean cardinalities of choice may be very close to each other, the asymptotic equivalence may be not true if the variance of these cardinalities is large.

The theorem can be proved with the use of the Chebyshev inequality.

Let us now discuss the results which establish the asymptotic equivalence of choice functions of various classes. Recall that a random subset in our case is a repeated independent sample of size $n \to \infty$ with some probabilistic measure on the total space of alternatives.

Before we verbalize the first theorem, consider one general result pertaining to the share of alternatives isolated by the heritage choice function from a random set. Namely, let $e(n)$ be a mean number of alternatives selected by the choice function from a random sample of size n. Then the share of the alternatives $e(n)/n$ chosen by heritage choice function does not grow with n.

The proof is straightforward, actually resting upon only the property of rearrangement of alternatives in a sample. Statistically, this proof reflects an informal similarity between the **H**-property and the condition of the inverse monotonicity of the choice function.

Theorem 1 (Fig. 1.7). Let the choice function C satisfy the **H**-property and the share of the alternatives chosen tends to $\alpha > 0$ with $n \to 0$. Then there exists a $B \in U$ such that the choice function C is asymptotically equivalent to threshold function C^B, and moreover, $\nu(B)$ (the measure of B) equals α.

The idea underlying the proof is as follows. Set B is a totality of points chosen with probability 1 in a random arbitrarily large context by choice function C. Giving such definition to B, we obtain a threshold function C^B. It is quite obvious that C^B almost surely lies within C, and that C^B is well-conditioned.

The only thing still left to be proved is that the ratio of mean numbers of alternatives chosen by C and C^B tends to 1 with $n \to \infty$. This is also almost obvious, since an alternative lying outside B is chosen with probability tending to zero in a large enough context.

What is actually meant in this theorem is more or less clear. If the choice is large enough and quite reasonable (that is, if condition **H** is satisfied), then practically the entire lot of alternatives may be grouped into two parts: "good" (those belonging to set B which we have somehow isolated out) and "bad," and we certainly choose the "good" ones (the quality check choice function).

Verbalize now the following *Theorem 2* (Fig. 1.7). Let choice function C be a union of a finite number of pair-dominant choice functions pairwise asymptotically independent of each other. Assume function C is well-conditioned. Then this function is asymptotically equivalent to the pair-dominant one.

The idea of the proof is as follows. First, let us construct binary relations R in the following manner. To specify a binary relation, it suffices to specify the upper cuts in all of its points. Take a point $x \in U$, and consider the upper cuts of binary relations in point x for those pair-dominant choice functions whose union yields the choice function C. These binary relations amount to a finite number. Take an upper cut with a minimal measure (or a minimal number, if there are several such cuts). This one will be exactly the upper cut in point x of binary relation R being constructed.

Let us find pair-dominant function C_R using this relation R. It is quite obvious that C_R lies inside function C almost certainly. It therefore remains only to estimate the ratio of the mean numbers of the chosen alternatives. To find the number of alternatives chosen by choice function C, the following equality is useful: $C = \bigcup_{i=1}^{k} C_i$, the C_i 's being pair-dominant. Hence

$$|C(X)| = \sum_{i=1}^{k} |C_i(X)| - \sum_{i,j} |C_i(X) \cap C_j(X)| + \sum_{i,j,l} |C_i(X) \cap C_j(X) \cap C_l(X)| - \cdots$$

$$+ (-1)^{k-1} \sum_{1,\ldots,k} |C_1(X) \cap \cdots \cap C_k(X)|$$

Note that the condition of the pairwise asymptotic independence of functions C_i and C_j for all i and j permits rather easy estimation of the mean value of that sum:

$$E\,|C(X)| \approx \sum_{i=1}^{k} E\,|C_i(X)|$$

Using again the asymptotic independence condition and an ad-hoc procedure of estimating the cardinality of the pair-dominant choice function one can show that

$$\sum_{i=1}^{k} E\,|C_i(X)| \approx E\,|C_R(X)|$$

Now the theorem is obvious since C is well-conditioned.

Example. Let set U be the $\mathbf{R}^2$ plane. Measure ν is uniform in the unit square and the choice function C is collected-extremal, i.e. it is a union of two choice functions C_1 and C_2 ($C = C_1 \cup C_2$) where C_i chooses an alternative with a maximal i-th coordinate.

One can easily check that C is well-conditioned (the dispersion of the choice cardinality is $O(1/n)$ and its mean tends to 2), and that C_1 and C_2 are asymptotically independent: the mean cardinality of the intersection of C_1 and C_2 is $1/n$ and the cardinality of each of them is 1.

Hence the collected-extremal choice function proves to be asymptotically equivalent to the pair-dominant one.

The theorem, in fact, describes a situation of splitting the total set of alternatives into parts each having its own structure of preferences: although we can estimate different perfumes offered in a perfume department from the viewpoint of alcohol content, we never do so provided the liquors department nearby offers a good choice.

We now pass over to the α-Tournament choice function. Recall that the α-Tournament function prescribes us to choose that α-share of the participants who have gained the largest score ($0 \le \alpha \le 1$) in the tournament. Recall also that, like the conventional tournament choice, the α-Tournament choice does not satisfy any rationality conditions and, in fact, may be highly irrational, since an addition or removal of a participant may dramatically change the outcome of the tournament. Nevertheless, from the viewpoint of the asymptotic equivalence, i.e. in large-scale random tournaments, the α-Tournament choice function works very well.

Let us give a rigorous formalization to these results.

Assume that f is a measurable and bounded score function. Let

$$\mu(x) = \int_U f(x,y)\,d\nu(y)$$

be the mean number of scores gained by team x in the games against a random opponent, and

$$\phi(\mu) = \nu(x \in U:\ \mu(x) < \mu)$$

is the share of alternatives for which this mean number of scores is less than μ.

Theorem 3 (Fig. 1.7). The α-Tournament choice function $C^{T\alpha}$ is asymptotically equivalent to the threshold function C^B where set B is comprised of the participants for whom the mean number of scores gained in their game against a random player is less than μ^*, where μ^* is a root of the equation $\phi(\mu) = \alpha$.

Set B actually includes the participants with best mean results.

The meaning of the theorem is thus in the fact that, under the given conditions (f being measurable and bounded, and ϕ continuous), the outcome of the tournament in terms of the scores comes close to the mean one.

For an individual participant this is obvious. The non-trivial fact about it is that the closeness to the mean is observed for all of them at the same time.

The theorem is proved according to the following general procedure: an auxiliary choice function $C = C^{T\alpha} \cap C^B$ is constructed which approximates each of the two functions from inside. Both are well-conditioned.

Simple probabilistic considerations using the Chebyshev inequality permit estimating the mean cardinality of choice by function C:

$$e_c(n) \approx \alpha n$$

Thus according to the first auxiliary theorem the choice function C is asymptotically equivalent to $C^{T\alpha}$ and C^B, and consequently, $C^{T\alpha}$ and C^B are asymptotically equivalent.

A natural question may be asked: in which cases the statistical conditions of the theorems, i.e. the independence and well-condition properties are satisfied. An answer to this question requires studying of stochastic characteristics of the choice functions specified by binary relations. In many cases we can answer this question when the choice function, for example, is constructed of binary relations defined in the critical space which are not too bad in some sense, or when these binary relations are defined on components of a direct product.

Conclusion

What is the meaning and essence of such theorems? (The author is sure that they are bound to grow in number involving other choice functions of theoretical and practical importance). By their very essence these results are devoid of any sensational paradoxicality. They do not state that something rational is impossible, or something irrational is possible. Rather, these results attempt to bring the theoretical understanding of rationality closer to the common sense removing their paradoxical incompatibility. A paradox is always a result of inadequacy of the theory or the language for describing the objective phenomenon. The existence of a paradox is indicative of either incompleteness or inaccuracy of the model. The possibility of removing some of such inadequacies within a model of choice on random sets shows, in the author's opinion, the vitality of the model.

REFERENCES

1. Baryshnikov, Yu.M. and Berezovskiy, B. A. On the asymptotic equivalency of choice functions. Avtomatica i Telemekhanika, No.10, 1986, pp.101-1-5 (in Russian).
2. Baryshnikov, Yu.M., Berezovskiy, B. A., Borsenko V.I., Kempner L.M. "Multicriterial optimization: Theory and applications." *Optimization Software*, New York (to appear in 1988).

STATISTICAL PROPERTIES OF CHOICE FUNCTIONS IN CRITERIAL SPACE

Introduction

A desirable goal of the theory of choice functions on random sets is to get an insight into the "physiology" of a choice function, that is, to find the way a choice function processes a random presentation.

This certainly requires recognition of both the structure of the choice function and the structure of the random subset which serves as its argument.

Let us again recall some basic definitions of the choice theory (Fig. 2.1).

We say that a choice function satisfies the heritage (**H**) condition if each alternative chosen in a larger subset will be inevitably chosen in a smaller one.

The condition of independence from rejected alternatives (**O**) implies that if an alternative is not chosen in a subset, then its removal from that subset doesn't effect the choice.

The extended concordance ($\mathbf{C}_k$) condition states that, for any k subsets $X_1, \ldots, X_k$, the fact that alternative x is chosen in each of them implies that a pair of subsets X_i, X_j exists such that alternative x is chosen in their union.

These conditions, possibly somewhat modified, make a kernel of the so-called rationality axioms. The satisfaction of these axioms is most preferable for the admission of the choice function as reasonable or rational.

Consider an important class of such functions: pair-dominant choice functions (Fig. 1.3, 1.4).

Let a binary relation R on the set U be given. Then a pair-dominant function C_R isolates exactly those alternatives x from the subset X for which there's no variant y in X such that $y\ Rx$. Relation R is usually interpreted as a preference or dominance relation while the alternatives chosen by function C_R are referred to as nondominated.

The role of a paramount importance played by pair-dominant functions in the choice theory is explained by the following facts.

First. If choice function C satisfies the **H**-condition, then there exists a set (family) of binary relations R_i such that the choice function C coinsides with a union of the corresponding pair-dominant choice functions C_{R_i}. Thus, a very natural **H**-condition guarantees that a choice function may be composed of pair-dominant ones. Note that the set of binary relations that determines these pair-dominant choice functions is not necessarily finite.

Second. If, besides the **H**-condition, the $\mathbf{C}_k$-condition is also satisfied in the choice function, then this set may be chosen finite with a cardinality of $k-1$.

Condition $\mathbf{C}_2$ coincides with a classical definition of concordance, and satisfaction of conditions **H** and $\mathbf{C}_2$(=**C**) at one time is equivalent to the coincidence of the choice function with the pair-dominant one in terms of some binary relation.

Third. Satisfaction of yet another condition, that of rejection, allows a transitive binary relation to be chosen.

Let us give a description of the basic characteristic that we shall deal with. As was noted above, the subject of this discussion is the effect of a deterministic choice function on random sets. It is evident that the random set's choice is again some random set. The cardinality of this set, designated S, is actually the basic value that we are going to investigate.

The following reasons have dictated the choice of this particular value:

(a) If in the general case the cardinality of the choice made in a classical optimization setting (i.e. in choosing an alternative which maximizes a given goal function) is unity then in case of an arbitrary choice function the cardinality of choice may vary. It is exactly the plurality of choice that strikes most in turning from the classical setting to a nonclassical one — for instance, to multicriterial.

(b) A general choice function, especially one for choosing from a random set, is an object which does not lend itself to straightforward understanding. Anyway, at first sight it may be difficult to say anything reasonable on the compared choice functions and find out which one is stronger and which is weaker. Yet the cardinality of choice is a number, even though a random one, and the knowledge of this value gives us certain quantitative characteristics of the choice which psychologically makes its acceptance easier.
Thus, in real-life multicriterial optimization one method (function) of choice is regarded to be better than another if it isolates a lesser number of alternatives; in other words, in order to compare choice functions they should be somehow measurable.

(c) And, finally, the third reason is in the fact that the knowledge of statistical characteristics of the choice function is most critical in other sections of the theory of choice on random sets. Thus, all theorems on asymptotic equivalence are in fact resting upon some statistical properties of choice functions. The same is true of the best choice problem.

We shall concentrate our attention at the study of the cardinality of choice over pair-dominant functions. The results of this study are easily extendable to the choice functions satisfying H- and C_k-conditions for $k > 2$ through the following reasoning: a choice function having these properties is a union of a finite number of the pair-dominant choice functions $C(X) = \bigcup_{i=1}^{k-1} C_{R_i}(X)$. Hence the cardinality $|C(X)| = |\bigcup_{i=1}^{k-1} C_{R_i}(X)|$ which according to the inclusion/rejection formula is

$$\sum_{i=1}^{k-1} |C_R(X)| - \sum_{i,j} |C_{R_i}(x) \cap C_{R_j}(X)| + \cdots + (-1)^{k-1} |\bigcap_{i=1}^{k-1} C_{R_i}(X)|.$$

It remains only to note that an intersection of pair-dominant choice functions is again a pair-dominant choice function determined by a binary relation—a union of the corresponding binary relations.

We shall mostly deal with random subsets which may be specified in the following manner (Fig. 2.2, 2.3). Let a Σ-algebra G on the set U, and a probabilistic measure ν be given; then we assume that X is a repeated independent sample from U of some size N which, generally speaking, is a random size. To put it differently, the sample is generated as follows: first we randomly choose a number N, then place N points $x_1, \ldots, x_N$ into U in accordance with probabilistic distribution ν. Let us make some more general remarks. In order to be able to say anything about the statistical characteristics of cardinality of our choice, this must be a measurable value.

Assertion. Let a binary relation $R \subseteq U \times U$ be measurable with respect to Σ-algebra $G \times G$. Then the cardinality of choice made according to the pair-dominant choice function is measurable, too.

A corollary follows. If $C = \cup C_{R_i}$ and all R_i's are measurable with respect to $G \times G$, then the cardinality of choice by C is measurable, too.

Combinatorial Statements and Results (Fig 2.4, 2.5)

The problem of studying statistical properties of choice cardinality stated somewhat generally has been verbalized a long time ago. Thus for instance, the problem of estimating the number of points on a random set's convex hull, which easily lends itself to a reformulation in choice theory terms, has been under discussion since the end of the last century. However, it seems more reasonable to start from another statement which is much closer to multicriterial optimization and choice theory problem.

Let $U = \mathbf{R}^n$, and let the coordinates of a random point be independent random values with continuous distribution function, i.e. measure ν, which is a direct product of continuous measures over the coordinates. Let, furthermore, binary relation R be Pareto-type, i.e. let one alternative dominate another if the former one is better in all coordinates.

What can be said about value $\mathrm{S}(n)$, the number of Pareto-nondominant choice alternative of size n? This problem has been first stated by O. Barndorff-Nielson and M. Sobel in 1966 in connection with some statistical applications. In this work, they have suggested an expression for a mean number of nondominant alternatives

$$E\,\mathrm{S}(n) = \sum_{1 \le i_1 \cdots \le i_{m-1} \le n} \frac{1}{i_1 i_2 \cdots i_{m-1}}$$

and established the asymptotics for this value $\approx \dfrac{ln^{m-1} n}{(m-1)!}$, where m is the number of coordinates. Besides, they have found dispersion $\mathrm{S}(n)$ with $m = 2{,}3$ (it turned out that the dispersion has the same order of growth as the mathematical expectation), and proved that with $m = 2$ distribution $\mathrm{S}(n)$ is asymptotically normal.

In later years, many of the results obtained by these authors have been rediscovered in various applications, such as estimation of the number of conventionally-optimal trajectories in dynamic programming, liberalistic paradox, etc. Different corrections have been made pertaining to this case: thus, a general formula for dispersion DS(n) has been found, and its asymptotics estimated for m=4. The asymptotic decomposition for the mathematical expectation has been corrected, and the asymptotics studied with $n = \text{const}$, $m \to \infty$ and $n/m = \text{const}$, m and n both $\to \infty$. Considered likewise was the case of the Pareto-type binary relation, but of another probabilistic measure. For instance, if the measure is multidimensional and normal, then the asymptotics of the mean number of nondominant alternatives as $n \to \infty$ is $c\ \ln^{m-1} n$, where c is a constant depending upon the covariance matrix. In the same way, consideration was given to a uniform measure in a unit simplex. In contrast to the previously obtained results, it turned out that the mean number of nondominant alternatives behaves as $n^{(m-1)/m}$ rather then as the power of the logarithm. These results used to be quite popular among multicriterial optimization people, in the USSR at least, for the purpose of apriori estimation of the strength of multicriterial techniques. Ignoring narrow statements of their problems, they thought the logarithm of the sample size to be a more or less universal law of the nondominant alternatives number growth.

Regardless of a great number of works in this area clearly indicative of its application validity, the problem foumulation was rather narrow, which may certainly be explained by limited capabilities of combinatorial techniques which seemed so natural in tackling such problems.

Distribution Function of the Binary Relation Upper Cut Measure (Fig. 2.6)

Consideration of more general formulations necessitated the development of some other techniques which will be described below.

Some definitions must first be made. Let $a \in U$. An upper cut of the binary relation R at the point a (designated R_a) is defined to be the set of elements $a' \in U$ such that $a' \in R_a$. The upper cut measure at point a, $\mu(a) = \nu(R_a)$ is a function measurable on U. Let x be a random point. Then value $\mu(x)$ is also random. Denote its distribution function through $B(\mu)$; in other words, $B(\mu)$ is a measure of the totality of such points on U whose upper cut measure does not exceed μ. It turns out that it is exactly the function $B(\mu)$ that determines the behavior of the mean number of nondominant alternatives. To put it more accurately, let $\phi(z)$ be the moment generality function of random value N, i.e.

$$\phi(z) = \sum_{n=1} P(N = n) z^n$$

Assumption. The mean number of nondominant alternatives in a sample of random size N is determined by the formula

$$\int_0^1 \frac{d\phi}{dz} (1 - \mu) dB(\mu).$$

Thus, the mean number of nondominant alternatives is a certain integral transformation of the distribution function for the random point's upper cut.

To prove the formula for an arbitrary N, it is sufficient to prove it for $N = n$ and then randomize the result. The proof of the formula in the above particular case rests upon the use of the following consideration: the mean number of nondominant alternatives equals the size of the sample times the probability of nondomination of a fixed element of the sample. This probability is the integral taken over the entire U with respect to measure ν of the measured function $(1 - \mu(a))^{n-1}$. Applying the Fubini theorem we obtain the above formula.

Consider two most important cases of a random value from a sample of size N.

Case 1: $N = n$. Then $E\,\mathrm{S(n)}$ is the Mellin transform of the distribution function $B(\mu)$.

Case 2: N is Poisson-distributed with parameter λ, i.e. the probability of $N = n$ is $\frac{\lambda^n}{n!} d^{-\lambda}$.

In this case, the mean number of nondominant alternatives is the Laplace-Stieltjes transformation of the distribution function $B(\mu)$.

Denote the mean number of nondominant alternatives as $e(n)$ in the first case and as $e(\lambda)$ in the second case.

This simple correlation between the mean number of nondominant alternatives and the distribution function of the upper cut measure permits the use of properties of the integral transformations for obtaining the asymptotic decomposition of function $B(\mu)$ as $\mu \to 0$, generating the asymptotic decomposition of $e(n)$ or $e(\lambda)$ as n and $\lambda \to \infty$.

The second important property is stated in the Tauber theorem: if $B(\mu)$ is a regularly changing function with $\mu \to 0$, i.e. if it may be presented as $\mu^{\alpha} h(\mu)$, where $\lim\limits_{\mu \to 0} \dfrac{h(t\mu)}{h(\mu)} = 1$ for any t, then $e(n)$ grows as $n^{1-\alpha} h\left(\dfrac{1}{n}\right)$ and $e(\lambda)$ grows as $\lambda^{1-\alpha} h\left(\dfrac{1}{\lambda}\right)$.

Consider a simple classical example illustrating the concepts introduced and the results obtained above.

Let set U be $\mathbf{R}^2$, measure ν be uniform on the unit square, and binary relation R be the Pareto relation. Then the measure of the upper cut in a point with coordinates (x, y) $(0 \le x \le 1, 0 \le y \le 1)$ is $(1-x)(1-y)$. The totality of points where the measure of the upper cut is less than μ is bounded by the sides of the square and by the hyperbola $(1-x)(1-y) = \mu$. The measure of this totality, coinciding with its area in this particular case, is $\mu(1 + \ln \frac{1}{\mu})$. Without any calculation, this directly leads to the well-known result:

$$e(n) \approx \ln n.$$

Let us switch to consideration of far more general results.

Mean of Nondominant Alternatives (Fig. 2.9)

As was noted above, our attention is paid mostly to the choice functions resulting from binary relations. The theory suggested below deals with even more concrete objects which are most natural and most often met in multicriterial optimization problems. To be specific, let us regard the alternative space U to be an m-dimensional real space $\mathbf{R}^m$ ("criterial space").

To be able to describe the measures and binary relations considered, we shall introduce a class of geometrical objects.

Definition. A locally-conic set in $\mathbf{R}^m$ is such a set V such that for any point x of this set, there exists some neighborhood u containing x, and a dipheomorphism $g: u \to u' \subset \mathbf{R}^m$, such that $g(x) = 0$, and $g(\mathbf{V} \cap u)$ is a cone in $\mathbf{R}^m$.

Let us now describe the class of measures ν considered on space $\mathbf{R}^m$. We shall regard measure ν to be specified by a continuous density carried by the locally-conic set $\mathbf{V}$. This density approaches the boundary of $\mathbf{V}$ in a power form, i.e. it may be presented as $b(x) h^s(x)$, where

$b(x) > 0$ for all $x \in \mathbf{V}$, b is continuous, $s > -1$ and $h(x)$ is the distance from point x to the exterior of $\mathbf{V}$. Note that we do not specify the metric in which distance $h(x)$ is measured, because all of them are equivalent (in a finite-dimensional space). Substitution of one metric for another may only change the form of function $b(x)$. These are the assumptions we make on measure.

Consider now the binary relation R. Assume this binary relation is specified by cone K; i.e. $x\ R\ y \Leftrightarrow x - y \in K$. This assumption is typical for multicriterial problems. We shall assume cone K coincides with the closure of its interior.

Note the following: any locally-conic set is stratifiable, i.e. it may be presented in the form of smooth cones-stratas-of different dimension. Introduce a notion of a weak optimum. Weak optimum is a set of points $\mathbf{V}$ in which the upper cut measure equals 0. The dimension of a weak optimum is the maximal dimension of the strata whose intersection with this optimum is not empty. Evidently, in the case of a solid cone this dimension does not exceed $m - 1$. We say that this dimension is stable if any small shifts of the cone do not change it; this means that cones K_1 lying inside K and K_2 containing K exist such that the dimension of the weak optimum over these cones coincide with that of the weak optimum over K.

Theorem. Let dimension K of the weak optimum be stable. Then the distribution function of the random point upper cut measure has the zero asymptotics of $\mu^{\frac{m-k+s}{m+s}}$. Hence, the asymptotics of $e(n)$ is

$$n^{\frac{k}{m+s}}.$$

The validity of this result is not quite obvious since its formulation is based upon a hard-to-check condition of the weak optimum dimension stability. Nevertheless, the following assertion is true (Fig. 2.9).

Consider group G of all possible rotations of space R^m and inside this space a subset Σ of rotations g such that for a measure $\mathbf{V}$ and cone gK the theorem's conditions are not satisfied, i.e. the condition of the weak optimum dimension stability is violated. Then the set Σ is actually contained in stratified subset g of dimension 1. In other words, if the conditions of the theorem are true for the pair ν and K, then they will be satisfied for any pair ν and K', where K' is obtained from K by means of a small rotation. On the other hand, if these conditions are not satisfied for some pair ν and K, then they may be made true with an arbitrarily small rotation of cone K. In such situation it is usually said that the conditions of the theorem are general-type conditions. The term "general-type conditions" means that they be regarded satisfied almost always, while violation of such conditions is indicative of some sort of specific constraints imposed upon the problem.

Summarizing the above, it may be noted that a typical multicriterial optimization problem features power asymptotics of the growth of the mean number of nondominant alternatives as a function of the sample size. This result is certainly a surprise, bearing in mind the logarithmic form in previous statements of the problem. (Note that the considered case of the uniform measure on the unit cube and the Pareto cone obviously belongs to the class of problems we are interested in.)

The conclusion that may be made here is quite evident: the case with independent coordinates is exactly that particular case whose small rotation allows us to obtain the general one. Consideration of the deformation of this particular case is most useful for understanding the general result.

Let us look upon Fig. 2.10. The center figure depicts a uniform measure distributed within a unit square, and a Pareto cone. The weak optimum in this case has dimension 1 and consists of the upper horizontal and righthand vertical sides of the square. The dimension of this weak optimum is not stable. To show this, apply a slight compression on the sides of the square as shown in the lower figure, and you will obtain the situation when the weak optimum dimension gets down to zero as it actually consists of a single vertex of the parallelogram. The situations depicted in the upper and lower figures are, on the contrary, stable and therefore in these cases the power asymptotics is realized as prescribed by the theorem, the exponent of the power depending upon the dimension of the weak optimum which is zero for the lower figure and one, for the upper figure. Recall that in the situation considered $m = 2$ and $s = 0$. Considering this and similar cases leads us to a hypothesis stating that for any family of deformations, the mean number of nondominant alternatives, as a function of the sample volume, may be decomposed into an asymptotic series of the following form:

$$e(n) = \sum_{l,k<m-1} a_{l,k} n^{m+s} \ln^k n + 0(1),$$

where coefficients $a_{l,k}$ are actually semi-algebraic functions of the deformation parameters.

Consider one more example typical for applications. It involves the case when measure ν in the criterial space is an image of a uniform measure in the criterial space (Fig. 2.11).

Let $U = R^2$, and measure β be an image of a uniform measure on a sphere in its linear projection into R^2. In this case it is quite evident that $s = -0.5$ and the dimension of the mean optimum for any convex cone K not containing straight lines will be unity. Therefore the mean number of nondominant alternatives grows here as $n^{2/3}$.

A little digression from our main topic is due: there are two "parallel" directions of investigation which come pretty close to the above approach, if not statement-wise, then at least in their results.

The first direction treats the problems on the mean number of versions on a convex hull of a finite random set. These problems are faced in numerous applications from biology to economy. Consider just a few typical results to underline the similarity with our situation.

If n points are uniformly thrown into a square, then the boundary of the convex hull will hold $\approx \ln n$ points. On the other hand, if the same n points are uniformly thrown into a circle, then the order of vertices will be $n^{1/3}$. A similar value for an m-dimensional sphere is

$$n^{\frac{m-1}{m+1}}.$$

The second direction deals with studying the asymptotics of fast-oscillating integrals which are of paramount importance in physical applications. The basic underlying fact is as follows: if f is a polynominal and g is a finite function, then

$$\int e^{-\tau f(x)} g(x) dx = \sum_{\alpha,k} a_{\alpha,k} \tau^{\alpha} \ln^{k},$$

where α runs through some rational arithmetic progression fully determined through f (compare with $\frac{l}{m+s}, l = 1, \ldots, m-1$, and l is an integer not exceeding the dimension of the space.

Dispersion and Distribution Function of Nondominant Alternatives (Fig. 2.12)

Let us now switch over from discussing the mean number of nondominant alternatives to further in-depth study of properties of value **S** for the class of problems we are interested in, namely the class of problems fully determined by a certain form of density in m and to the binary relations that are specified by cone K.

To obtain more accurate information pertaining **S**, we certainly have to pay the cost of the more stringent requirements imposed upon the conditions of the problem.

Satisfaction of the following condition is assumed below: any small shift of cone K has no effect on the weak optimum.

This condition, generally speaking, is not the general-type condition. To show this, imagine a sphere: it is evident that a small shift of the cone does change (although insignificantly) the weak optimum.

Note, though, that there are important (and wide enough) classes of density supports for which this condition is always satisfied, for example, the class of convex polyhedrons (in case when K is convex, too.)

Note, besides, that the theorem given below, most probably does not require this more stringent condition. However, avoiding this condition is matter of further studies.

Theorem. Under the assumptions made and if cone K is convex and the number of nondominant alternatives tends to infinity with a growth of the sample volume (or, which is the same, if the weak optimum dimension is greater than zero), dispersion **S** increases as a mean and the centered and normed **S** is asymptotically normal.

Note the parallel between this result and the one presented above concerning the number of Pareto-nondominant alternatives in the case of two independent criteria where the dispersion grows as a mean as well, and where the distribution is asymptotically normal. Unfortunately, today we can give no other examples which could verify the hypotheses that suggest themselves in this case.

The above theorem satisfies almost fully the practical need in the information on the behavior of the mean number of nondominant alternatives. Indeed, according to this theorem the distribution **S** taken over large samples is almost fully determined by two values: its mean and its dispersion. These values may be easily calculated—if not analytically, then with the use of numerical modelling techniques (Monte-Carlo method).

If we recall that in the majority of applications this density actually meets our assumptions (the reason being that, as a rule, this density is an image, with a smooth criterial mapping, of some plane on the set of feasible criteria which itself is rather well designed), then a conclusion may be made that the above theorems applied to problems with conical relations do offer a full description of the value **S**.

Special cases (Fig. 2.13, 2.14)

However, there are numerous problems of finding the properties of **S** for various special cases of binary relations snd probabilistic measures. Treating these cases may be explained by two reasons: either they are traditional statements, and their study is a bow of respect to those who have pioneered them, or they appear in an investigation of a theoretical question. These special cases are as follows.

Direct product. Assume that total space, binary relation and probabilistic measure are decomposable into direct products. A typical example of such a case is Pareto comparison in R^m (which is a direct product of m linear orders in $\mathbf{R}^1$) with independent criteria (i.e. the probabilistic measure is a product of measures on the cofactors). A natural and important question arises, as to whether we can say anything about the statistical characteristics of **S** knowing similar values in the cofactors. If we are interested in the mean value of S as a function of the sample volume, then the answer to this question is yes.

Theorem. The mean number of nondominant alternatives in a sample of size n is determined by mean numbers of nondominant alternatives in the cofactors in samples of a size less than or equal to n.

A precise formula is rather cumbersome but we can easily understand its intuition: as was noted above, the knowledge of function $e(n)$ is about the same as the knowledge of the distribution function for the upper cut measure. And it functions $F(\mu)$, $F_1(\mu)$ and $F_2(\mu)$ are the distribution functions for the product and its cofactors, respectively, than the following formula is true: $G(\eta) = G_1(\eta) * G_2(\eta)$, where $G_i(\eta) = F_i(e^\eta)$ and "$*$" is the convolution.

A qualitative description of the behavior of the operator which converts the pair of functions $e_i(n)$ coresponding to the cofactors into functions $e(n)$ corresponding to their product may be given as the follows: if $e_i(n)$ have different orders of growth, then $e(n)$ is equivalent to a larger one, while if $e_1(n) \approx e_2(n)$ are changing slowly, then $e(n) \approx e_i(n) - \ln n$. This explains the way m units $e_i = 1$ are "glued together" in the classical problem to make $\ln^{m-1} n$.

It is interesting to note that the dispersion and higher-order moments in the product are not determined by the corresponding and lesser-order moments in the cofactors.

It would be apt to finish this lecture with a description of one problem which is obviously related to our topic but features an essentially different type of presentation (Fig 2.15).

Let a total space U be countable, and binary relation R in this space be given. Assume that random sample X is obtained in the following manner: for any $a_1, a_2 \in U$, events $\{a_1 \in X\}$ and $\{a_2 \in X\}$ are independent. Then, if $\sum_{a \in U} P(a \in X) = \infty$, we say that with probability 1 the cardinality of X is infinite. In the problem statements described earlier the infinity of the sample has lead to the emptiness of the set of nondominant alternatives. In our case, provided some natural conditions are met (like the condition of the finiteness of the upper cut for the "majority" of the points), this set is nonempty and possesses some very interesting specific features. Many examples may be cited when, with probability 1, the set of nondominant alternatives is finite, and its mean is finite too. In the particular case when all $P(a \in X) \equiv p$, the formulas used for the mean number of nondominant alternatives found as a function of p look very much like the formulas of "similar" binary relation in $\mathbf{R}^m$.

REFERENCES

1. Barndorff-Nielson, O., Sobel, M. "On the distribution of the number of admissible points in a vector random sample." *Probability theory and its applications*, Moscow, 1966; Vol. XI, no. 2, pp. 283-305.
2. Berezovsky, B. A., Borzenko, V. I., Kempner L.M. "Binary Relation in Multicriterial Optimization." *Nauka*, Moscow, 1981 (in Russian).
3. Baryshnikov, Yu. M. "On the mean number of options, non-interior in a binary relation." *Automatica i Telemekhanika*, Moscow, 1985, no. 6, pp. 111-116 (in Russian).

THE BEST CHOICE PROBLEM: OLD AND NEW

1. Introduction

The topic of this lecture is another class of problems related to the choice theory, which are traditionally named the best choice problems. As for their mathematical nature, these problems are problems of optimal stopping theory and from a general point of view are problems of statistical sequential analysis. Any problem of this type presents a situation where we deal with a sequence of stages and at each stage we have to accept a decision which influences the future process. Naturally, we want to accept such a decision which turns out to be optimal, but we must take into account that a decision which seems to be good at the present may land in a mess in future, leaving us to choose a decision that looks less attractive at the time. In any optimal stopping problem we have only two alternatives at a stage: we can stop observation process or continue. There is extensive literature on problems of best choice (see references in [1]). The interest in these problems is twofold: first, they reflect some essential aspects of real-life decision processes and, second, they always have meaningful statements and their solutions may be easily interpreted. In the best choice problems selection is performed as a sequential process of comparing and examining variants, there being some strategic and informational restrictions imposed upon the selection process. These restrictions result from unavailability of rejected variants and from strategic uncertainties in the quality of the forthcoming variants. The efficiency of selection depends upon comparisons of the chosen variant with others and increases with its quality.

In the classical statement we can arrange all the variants comparing them with respect to a certain criterion, the unit rank variant being considered as the best. For the optimality criterion of a choice strategy (in the classical problem these are stopping rules) the probability of choosing the best variant is taken. The literature on this problem is mostly devoted to the case of linear ordering of variants and the optimality criterion is the mean of a function of the rank of the chosen variant (e.g., the mean rank [2]).

In this lecture we give the statement and the solution of the classical problem and then consider two examples with non-classical choice functions. Generalizing these examples, we obtain a general statement of the problem. Dealing with optimality of the so-called threshold stopping rules, we find close relations between abstract set-theoretical rationality properties of choice functions and their probabilistic properties. In the general statement we reject the use of traditional dynamic programming methods, as their realization requires exponentially growing number of calculations, when the number of variants grows (a typical example is the Pareto choice function). Instead of that, we introduce different classes of suboptimal stopping rules, studying their efficiency for some choice functions.

2. *The Classical Statement of the Best Choice Problem*

It seems that the first statement of the optimal stopping problem was proposed by A. Cayley more than a century ago, and one can see his formulation in the book [10]. As for the classical best choice problem, its author is still unknown, though F. Mosteller [3] reports that he had heard about it in 1955 from A. Gleason, who in turn referred to somebody else. In the early sixties the problem became popular and was published in some magazines as a puzzle, for example, in M. Gardner's column of Scientific American [4] as a game named "googol," which consists in guessing the maximum number in a random sequence. By the middle sixties, the problem with numerous extensions passed to professionals, and now presents an inexhaustible source of examples for many sequential decision methods. It also has popular names such as "The secretary problem" or "The marriage problem".

Suppose we have N variants ordered by some quality sign and we want to choose the best one. We see the variants at random, every possible permutation being equiprobable. At each stage, we may compare the variant under consideration with its predecessors, but we know nothing about the quality of the future variants. Depending on the known comparisons, the next variant may be accepted (in this case the whole process terminates) or rejected (and in this case we see one more variant, if there are still unknown variants). The problem is to maximize the probability of choosing the best variant.

In the context of the secretary problem variants are interpreted as girls who want to get the secretary job, each girl being accepted or rejected immediately after the interview.

For better understanding of what is meant by a choice strategy or a stopping rule, let us consider the simplest ones. We may always accept, for example, the first variant. By the equiprobability assumption, this variant is the best with probability $1/N$.

A more complicated rule, which prescribes to reject $N/2$ variants and then to choose the first variant better than all its predecessors, is more efficient. Namely, if the second best variant appears in the first half of the sample and the first best variant appears in the second half, then this rule will surely select the best variant. Hence, the probability of a sucessful choice is not less than 1/4.

To describe all the admissible stopping rules, we introduce random values reffered to as absolute and relative ranks. Let $x_1, \ldots, x_N$ be a random sequence of variants which are observed at moments $1, 2, \ldots, N$ respectively. The linear order hypothesis allows us to arrange variants in such a way that better variants have lower ranks. Define X_n, $n = 1, \ldots, N$, as the rank of x_n among $x_1, \ldots, x_N$ and call X_n the absolute rank. By the analogy, define relative rank Y_n as the rank of x_n among $x_1, \ldots, x_n$. The sequence $X_1, \ldots, X_N$ unambiguously characterizes the results of comparing $x_1, \ldots, x_N$. So absolute rank 1 corresponds to the best variant, absolute rank 2 corresponds to the second best variant and so on. The set of absolute ranks $(X_1, \ldots, X_N)$ is a random permutation of $1, \ldots, N$, all these permutations being equiprobable.

Taking as a basis of a probabilistic model the sequence of absolute ranks, we can easily find the mutual distribution of the sequence of relative ranks. Note that Y_n is the rank of x_n among $x_1, \ldots, x_n$ (e.g. $Y_n = 1$ if $X_n = \min(X_1, \ldots, X_n)$), so its values are $1, \ldots, n$. By the equiprobability of all the permutations, one can derive that $Y_1, \ldots, Y_N$ are independent and have uniform distribution, i.e. Y_n equals k, $k \leq n$, with probability $1/n$ independently from the values of other relative ranks.

The set $Y_1, \ldots, Y_n$ unambiguously identifies all the comparisons of the variants $x_1, \ldots, x_n$. Indeed, Y_k is the number of variants among $x_1, \ldots, x_k$, which are better than x_k. The opposite proposition is also true, as by the $Y_1, \ldots, Y_n$ we can reconstruct the comparisons of $x_1, \ldots, x_n$. For example, if $Y_n < Y_{n-1}$, then x_n is better than x_{n-1}. Thus, the values of $Y_1, \ldots, Y_n$ contain all the information available at the stage n. This means, that selection of the n-th variant must depend exclusively on $Y_1, \ldots, Y_n$.

So, we come to the following definition (Fig. 3.3). A stopping rule t is a random variable with values $1, \ldots, N$ which has the property that t equals n depending only on the values of $Y_1, \ldots, Y_n$, $n \leq N$. In the case $t = n$ we say that the rule t selects the variant x_n.

For example, the prescription "choose the first variant" is a stopping rule, but the prescription "pass all the variants and return to the best one" is not a stopping rule, because selection of x_n depends on X_n and the latter essentially depends on the whole sequence $Y_1, \ldots, Y_N$.

The efficiency of the stopping rule t is characterized by the formula

$$E(t) = \mathbf{P}\{X_t = 1\} \tag{1}$$

i.e., it equals the probability of stopping on the best variant. The sequence $Y_1, \ldots, Y_N$ takes on only a finite number of values, thus the number of all stopping rules is finite too, and hence there is an optimal stopping rule t^*, maximizing (1).

Regard x_n as a relatively best variant if $Y_n = 1$. It is clear that the last relatively best variant is the absolutely best one, so our problem is to recognize the last relatively best variant at the moment when it appears.

Suppose that the variants $x_1, \ldots, x_n$ are rejected, then at n-th stage we have a dilemma: we may choose x_n and stop the process, or we take the variant x_{n+1}. If the variant $x_n, n < N$, is not relatively best, the choice is not optimal since the selection of the following variant has non-zero probability of success. If the variant x_n happens to be relatively best, the choice is successful with probability

$$\mathbf{P}\{X_n = 1 \mid Y_n = 1\} = n/N$$

and we have to compare it with the probability of the successful choice which we get with optimal continuation of the selection process. Denoting the latter probability by v_n, if $n/N < v_n$, then it is reasonable to accept it. This idea is a particular realization of the general dynamic programming principle of Bellman, and it leads to the determination of the optimal stopping rule. By the independence of absolute ranks the value v_n is constant, i.e., it does not depend on $Y_1, \ldots, Y_N$. On the other hand, the sequence $v_1, \ldots, v_N$ does not increase. Indeed, when $x_1, \ldots, x_n$ are rejected, we can make the successful choice with probability v_n, but this is a way to pass $x_1, \ldots, x_{n-1}$, so $v_n \leq v_{n-1}$. On the contrary, the sequence $1/N, 2/N, \ldots, N/N$ is increasing, therefore inequality $n/N \geq v_n$ holds for all $n \geq d^*$ for some d^*. Hence, the optimal stopping rule is

$$t^* = \min\{n \mid n \geq d^*, Y_n = 1\}$$

where $\min \theta \underset{\text{def}}{=} N$.

We proved that the optimal choice strategy prescribes to reject $x_1, \ldots, x_{d^*-1}$ and then to accept the first relatively best variant. The number d* is the threshold which divides the selection process into two periods: during the first period we collect information and form the pattern, and during the second period we compare the variants with the pattern and select one of them if it is better.

There are different ways to compute $E(t^*)$. One of them is to write a recursion for $v_1, \ldots, v_N$ and find the solution, but an easier way is the following. Let

$$t_d = \min\{n \mid n \geq d, Y_n = 1\} \tag{2}$$

be a threshold stopping rule with a threshold d. From (1), (2) and the equality

$$E(t^*) = E(t_{d^*}) = \max E(t_d),$$

we conclude that

$$E(t_d) = \frac{d-1}{N} \sum_{k=d}^{N} \frac{1}{k-1} . \tag{3}$$

It follows that the optimal threshold d^* may be determined from the inequalities

$$\sum_{k=d^*+1}^{N} \frac{1}{k-1} \leq 1 \leq \sum_{k=d^*}^{N} \frac{1}{k-1} . \tag{4}$$

Solution of (4) allows us to compute d^* and v^*, i.e., for $N = 10$ we have $d^* = 4$ and $E(t^*) = 0{,}399$.

It is interesting to find the limit value of $E(t^*)$ when $N \to \infty$. To do this, you have to approximate (4) by the equation

$$\int_{d^*}^{N} dx/x = 1,$$

and then

$$\lim_{N \to \infty} d^*/N = \lim_{N \to \infty} E(t^*) = e^{-1} = 0{,}368.$$

We see that when the number of variants grows to infinity, the duration of the first period is near N/e and the optimal probability of the successful choice tends to $1/e$.

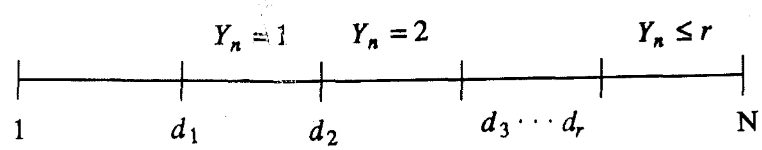
3. The Gusein-Zade's Problem (Fig. 3.4)

In the classical statement, the second best variant and the worst one are indistinguishable in the sense that their choice is equally unsatisfactory. In many real-life situations the choice of one of the r best variants proves to be more adequate, since the distinction among these r variants is not quite essential as compared to the increasing probability of successful choice. The formulation of this kind was proposed by Gusein-Zade [5], and for the case $r = 2$ by Gilbert and Mosteller. A detailed study of the case of large r is the subject of the paper [6].

Let $X_1, \ldots, X_N$ and $Y_1, \ldots, Y_N$ be the sequences of absolute and relative ranks as in the classical problem. In Gusein-Zade's problem for the stopping rule optimality criterion, the probability

$$E(t) = \mathbf{P}\{X_t \leq r\}$$

of selecting one of the r best variants is taken. In view of the inequality $X_n \geq Y_n$, the optimal stopping rule rejects all the variants whose relative rank is more that r. If $Y_n \leq r$ then the decision depends upon the comparison of the successful choice probability $\mathbf{P}\{X_n \leq r \mid Y_n\}$ with v_n, i.e., the gain from the optimal continuation. By the analogy with the classical problem, the optimal stopping rule is as follows: reject the variants $x_1, \ldots, x_{d_1-1}$, accept the first relatively best variant from $x_{d_1}, \ldots, x_{d_2-1}$, and if there is no such a variant, then accept the first variant from $x_{d_2}, \ldots, x_{d_3-1}$ whose relative rank does not exceed 2, . . . , if $x_1, \ldots, x_{d_r-1}$ are rejected, then accept the first variant from $x_{d_r}, \ldots, x_N$ whole relative rank does not exceed r, and finally if $x_1, \ldots, x_{N-1}$ are rejected then accept x_N. The choice process is schematically illustrated by the figure



where $d_1, \ldots, d_r$ are to be determined.

In principle, the optimal stopping rule may be explicitly determined by solving a recursive equation on v_n, or by maximizing $E(t)$ as a function of $d_1, \ldots, d_r$. The two procedures are rather tedious and we omit them, pointing out only that when $r = 2$, the limiting (when $N \to \infty$) value of $E(t^*)$ equals 0.573 . . ., value of $E(t^*)$ equals 0.573 . . ., and when $r = 10$ it already equals 0.976.

Instead of determining the optimal stopping rule, consider a simple extension of threshold stopping rules from the classical problem (2). Set up

$$t_d = \min\{n \mid n \geq d, Y_n \leq r\}$$

and call t_d a threshold stopping rule too. The rule t_d prescribes to reject the first $d-1$ variants and then to accept the first one whose relative rank does not exceed r. An easy calculation gives the successful choice probability formula, which generalizes the classical one:

$$E(t_d) = \frac{r}{N} \binom{d-1}{r} \sum_{n=d}^{N} \binom{n-1}{r}^{-1} \tag{6}$$

where $\binom{d-1}{r}$ is the binomial coefficient. Approximating the sum in (6) by an integral, we obtain the limiting relations for $r > 1$

$$\lim_{N\to\infty} E(t_{d^*}) = \lim_{N\to\infty} d^*/N = r^{-1/(r-1)} \tag{7}$$

For $r \to \infty$ the right hand side of (7) tends to 1, hence the class of threshold stopping rules becomes suboptimal (when $r = 10$ the right hand side of (7) equals 0,774 . . . and for $r = 100$ it equals 0,955).

4. Collective Extremal Choice (Fig. 3.5)

The following formulation is the simplest in the family of multicriteria best choice problems, in which the quality of a variant is characterized by evaluation in terms of a number of criteria. The detailed study is the subject of the paper [7].

Assume that the variants $x_1, \ldots, x_N$ are compared with each other in terms of a number of independent criteria, and each variant relates to a vector absolute rank. Any criterion relates to a linear ordering of $x_1, \ldots, x_N$ and we can associate with a variant x_n its absolute rank X_n in terms of k-th criterion, $k = 1, \ldots, m$. Thus, the absolute rank X_n is a column vector with components $X_n^1, \ldots, X_n^m$. It is natural to formalize the independence hypothesis as equiprobability of all $N!^m$ possible values of the sequence $X_1, \ldots, X_N$.

By the analogy with the classical problem, introduce the sequence $Y_1, \ldots, Y_N$ of relative ranks of $x_1, \ldots, x_N$ in terms of m criteria and consider stopping rules based on relative ranks. As stopping rule efficiency estimation we take the probability of stopping on one of the variants, which are the best in terms of some criterion

$$E(t) = \mathbf{P}\{(X_t^1 = 1) \vee (X_t^2 = 1) \vee \cdots \vee (X_t^m = 1)\},$$

we omit the case $m = 1$, corresponding to the classical problem.

To find the optimal stopping rule, consider the situation when $x_1, \ldots, x_{n-1}$ are already rejected and we decide if to accept x_n. It is clear that if $Y_n^k > 1$ for all k then there is no reason to accept x_n. If there are p units among $Y_n^1, \ldots, Y_n^m$ then the variant x_n remains the absolutely best in terms of one of the criteria with the probability

$$\mathbf{P}\{\bigvee_k (X_n^k = 1) \mid \vec{Y}_n\} = 1 - (1 - n/N)^p$$

As before, the choice decision must depend upon comparison of this probability with the gain v_n from the optimal continuation v_n. The optimal stopping rule is the following: reject $x_1, \ldots, x_{d_1-1}$ and then accept the first variant which has the unit relative ranks. If there is no such variant among $x_{d_1}, \ldots, x_{d_2-1}$ then accept the first variant which has $m-1$ unit absolute ranks. If there is no such variant among $x_{d_2}, \ldots, x_{d_3-1}$ then.... If the variants $x_1, \ldots, x_{d_m-1}$ are rejected, then accept the first variant which has at least one unit relative rank.

The optimal stopping rule may be explicitly determined by maximization of $E(t)$ as a function of $d_1, \ldots, d_m$, but this procedure provides poor information about asymptotics when $N \to \infty$. The limiting values are easily obtained by the following reasoning. First note that the optimal value of the threshold d_1 grows to infinity when $N \to \infty$ (else $v_0 \to 0$) and then estimate the probability of appearing among $x_{d_1}, \ldots, x_N$ such a variant which has at least two unit relative ranks. As far as $Y_n^{k_1} = Y_n^{k_2} = 1$, $k_1 = k_2$ with probability $1/n^2$ we have an estimation

$$\sum_{n=d_1}^{N} 1/n^2 \to 0 \text{ with } N \to \infty$$

It follows that since for large N, only the value of d_m is essential, a stopping rule of the form

$$t_d = \min \{n \mid n \geq d, \bigvee_k (Y_n^k = 1)\}$$

is asymptotically optimal. As regards the efficiency of threshold stopping rules t_d of the above type we have the known formula

$$E(t^*) \to E(t_{d^*}) \to \frac{m}{N} \binom{d-1}{m} \sum_{n=d}^{N} \binom{n-1}{m}^{-1}. \tag{8}$$

Finally, maximizing (8) we obtain

$$\lim E(t_d^*) = \lim d^* / N = m^{-1/(m-1)}. \tag{9}$$

5. The General Statement

Coincidence of the formulae (6), (7) with (8), (9), respectively, indicates that there exist some general facts on the efficiency of threshold stopping rules. The variables r and m included in these formulae have different definitions: the former is the maximal absolute rank of a "good" variant and the latter is the number of criteria in a multicriteria problem. For $N > r$ in the Gusein-Zade's problem we qualify r variants as the best, choice of the remaining variants being equally unsatisfactory. Similarly, in the second example when $N \to r$ the number of the "best" variants, which are optimizing one of the criteria, tends to m. Thus the variables r and m in the two problems may be interpreted as a number of variants among $x_1, \ldots, x_N$, which are qualified as the "best." There is a reason to relate the efficiency of threshold stopping rules with the number of the best variants. This idea leads to using the general concept of choice function in the statement of the best choice problem.

Let U be a total set of variants and $x_1, \ldots, x_N$ be a sequential sample from U such that the random variables x_n are permutable (i.e. the distribution of a vector $(x_{i_1}, \ldots, x_{i_N})$ is the same for any permutation $(i_1, \ldots, i_N)$). We consider $x_1, \ldots, x_N$ as a sequence of observable variants and want to select one of them, as before.

Assume that C is a choice function on U, this means that for a finite set $X \subset U$ a set $C(X)$ of the best variants is specified. Concider stopping rules based on the sequence $x_1, \ldots, x_N$. A variant x_n is defined as absolutely best if it belongs to $C(x_1, \ldots, x_N)$. In the same fashion, define x_n as a relatively best variant if it belongs to $C(x_1, \ldots, x_n)$. As efficiency criterion for a stopping rule t, we accept the probability of stopping on one of the absolutely best variants

$$E(t) = \mathbf{P}\{x_t \in C(x_1, \ldots, x_N)\}.$$

Introduce by analogy with (2) and (5) a threshold stopping rule

$$t_d = \min\{n \mid n \geq d, x_n \in C(x_1, \ldots, x_n)\},$$

which prescribes to reject the first $d-1$ variants and then to stop on the first relatively best one.

There is no way to obtain non-trivial lower-bound estimation of the threshold stopping rule's efficiency for all choice functions at once, even if we are restricted by that of choice functions, which satisfy $S_n = card\ C(x_1, \ldots, x_n) = \text{const}$, for all $n > n_0$. The necessity of restrictions is a matter of the fact that in general there is no connection between absolutely and relatively best variants, though only such a connection can ensure that relatively best variant remains absolutely best. In the case where a reasonable restriction is the condition H, we obtain as an implication

$$x_n \in C(x_1, \ldots, x_k) \rightarrow x_n \in C(x_1, \ldots, x_n), k > n.$$

Hence, an absolutely best variant is the relatively best variant at the moment of its arriving. Generalizing the previous results we have the following statements.

Theorem 1. Let a choice function C satisfy the condition **H** and for some $r \geq 1$

$$\mathbf{P}\{S_n = r\} = 1, n \geq r.$$

Then the efficiency of any threshold stopping rule t_d is expressed by the formulae (6) and (7).

Introduce one more condition of the choice rationality. We say that C satisfies the condition **O** iff

$$x_{n+1} \bar{\in} C(X_1, \ldots, x_{n+1}) \rightarrow C(x_1, \ldots, x_{n+1}) = C(x_1, \ldots, x_n)$$

Theorem 2. If C satisfies the conditions **H** and **O** and $\mathbf{E}S_n = r$ (here $\mathbf{E}S_n$ is the mathematical expectation of S_n), $S_n \leq m$ for all $n \leq N$ then

$$\max_d E(t_d) > \begin{cases} \frac{r}{m} m^{-1/(m-1)}, & \text{if } m > 1 \\ re^{-1}, & \text{if } m = 1 \ . \end{cases}$$

One can find the proof of these theorems in [9]. They are essentially based upon the choice function properties and permutability of $x_1, \ldots, x_N$.

6. Pareto Optimality (Fig. 3.8)

For a wide class of choice functions estimation of Theorem 2 is rather difficult. These are the cases when the number of the best variants S_n may take high values and the mean $\mathbf{E}S_N$ has a lower order. Nevertheless in some cases this estimate may be improved, as in the case of the Pareto optimality example.

Let $X_1, \ldots, X_N$ be the m-component vectors of absolute ranks of the variants $x_1, \ldots, x_N$, different components being mutually independent. Let $X_k > X_n$ if all the components of X_k do not exceed the corresponding components of X_n. A variant x_n is defined as Pareto optimal among $x_1, \ldots, x_N$, written $x_n \in C_{Par}(x_1, \ldots, x_N)$, if X_n is an undominated vector among $X_1, \ldots, X_N$ with respect to the partial order $>$. The variable S_n takes values from 1 to n and $\mathbf{E}S_n$ is of the order $\ln^{m-1} N/(m-1)!$, $m > 1$. Hence for the function of Pareto optimal choice,

$$\mathbf{E}S_n / \max S_n \rightarrow 0 \ \text{ when } n \rightarrow \infty.$$

To estimate the efficiency of the threshold stopping rules, define the variables

$$\pi_d = \text{card}\ \{n \mid d \leq n \leq N, x_n \in C_{Par}(x_1, \ldots, x_N)\}$$

and

$$\mu_d = \text{card}\ \{n \mid d \leq n \leq N, x_n \in C_{Par}(x_1, \ldots, x_{d-1}, x_n)\}.$$

One can prove that the function of Pareto optimal choice satisfies the conditions H and O and derive that the first variant $x_n, n \geq d$, which satisfies $x_n \in C_{Par}(x_1, \ldots, x_{d-1}, x_n)$ is Pareto optimal among $x_1, \ldots, x_n$. It follows that if we can consider threshold stopping rules of the form

$$t_d = \min\ \{n \mid n \geq d, x_n \in C_{Par}(x_1, \ldots, x_{d-1}, x_n)\},$$

for such a stoppping rule and a choice function satisfying the condition H the following formula is true

$$E(t_d) = \mathbf{E}(\pi_d / \mu_d)\,, \tag{10}$$

where $0/0 \underset{def}{=} 0$. Estimating the mean (10) by the Cauchy - Schwartz inequality, we can obtain the asymptotic value of the threshold stopping rule's efficiency for Pareto optimal choice with independent criteria [8,9]. It follows that

$$\max_{d} E(t_d) \to \text{,when } N \to \infty.$$

i.e. the optimal threshold stopping rule allows us to select a Pareto optimal variant almost with probability one when the size of the sample $x_1, \ldots, x_N$ is great enough. The latter result may also be extended to the case of dependent criteria. To do that, we have to use asymptotic results of the preceeding lectures on the probabilistic behaviour of S_n when $n \to \infty$.

REFERENCES

1. Berezovskiy, B. A. and Gnedin, A. V. *The best choice problem.* Moscow, Nauka, 1984 .
2. Chow, Y.S., et al. "Optimum selection based on relative rank (the 'secretary problem')." *Isr. J. Math.* 1964, v.2. No. 1, p. 81-90.
3. Gilbert, J. and Mosteller, F. "Recognizing the maximum of a sequence." *J. Amer. Stat. Assoc.*, 1966, v.61, No.313, p. 35-73.
4. Gardner, M. "Mathematical games." *Scient. Amer.*, March 1960, p.178-179.
5. Gusein-Zade, S. "The choice problem and optimal stopping rule of a sequence of independent trials." *Theory Prob. Appl.* 1966, v. 11, No. 3.
6. Frank, A. and Samuels, S. "On an optimal stopping problem of Gusein-Zade." *Stochast. Proc. Appl.*, 1980, v. 10. No. 3, p. 299-311.
7. Gnedin, A. "Multicriteria optimal choice problem." *Automation and Remote Control*, 1981, No. 7.
8. Gnedin, A. Effective stopping on a Pareto optimal variant. Ibid., 1983, No. 3.
9. Berezovskiy, B. A., et al. "On a class of best choice problems." *Inform. Sci.*, 1986, v. 39, p. 111-127.
10. Dynkin, E. and Yushkevich, A. *Markov processes: Theorems and problems.* Moscow, Nauka, 1967.

a. CLASSICAL MODEL OF CHOICE

(CHOICE FROM DETERMINISTIC SETS)

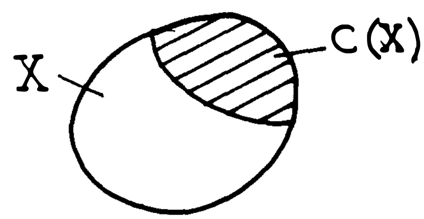
$\mathfrak{A}$ – INITIAL SET OF ALTERNATIVES (VARIANTS)

$\mathfrak{X} \subset 2^{\mathfrak{A}}$ – SET OF ADMISSIBLE PRESENTATIONS

$X \in \mathfrak{X}$ – PRESENTATION

$$C : \mathfrak{X} \to 2^{\mathfrak{A}}, \text{ WHERE } \forall X \in \mathfrak{X} \quad C(X) \subseteq X$$

$C(X)$ – CHOICE FROM X



b. EXTENDED MODEL OF CHOICE

(CHOICE FROM RANDOM SETS)

$$\begin{tikzcd} \text{MAN} \arrow[dr] & & \text{NATURE} \arrow[dl] \\ & C(X) & \end{tikzcd}$$

X IS A RANDOM PRESENTATION

FIG. 1.1

EXAMPLES
OF WIDE SPREAD CHOICE FUNCTIONS

1. SCALAR - EXTREME CHOICE FUNCTION

$$f : \mathfrak{A} \to \mathbb{R}$$

$$C^{E}(X) = \underset{x \in X}{\mathrm{ARGMAX}}\, f(x)$$

2. PARETO CHOICE FUNCTION

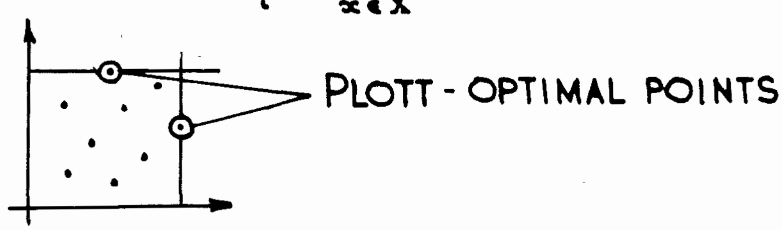
$$\mathfrak{A} \subset \mathbb{R}^{n}$$

$$C^{PAR}(X) = \{ x \in X : \forall y \in X \ \exists i : \ y_i < x_i \}$$

3. PLOTT CHOICE FUNCTION (COLLECTED EXTREMAL)

$$\mathfrak{A} \subset \mathbb{R}^{n}$$

$$C^{PL}(X) = \bigcup_{i} \underset{x \in X}{\mathrm{ARGMAX}}\ x_i$$



4. TOURNAMENT CHOICE FUNCTION

$$f : \mathfrak{A} \times \mathfrak{A} \to \mathbb{R} \quad \text{(FUNCTION OF SCORES)}$$

$$X = \{ x_1, \dots, x_k \}$$

$$m_X(x_i) = \sum_{j} f(x_i, x_j)$$

$$C^{T}(X) = \underset{x \in X}{\mathrm{ARGMAX}}\ m_X(x)$$

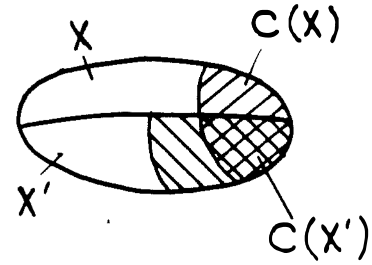
$$C^{T\alpha}(X) = \{ x \in X : \mathrm{card}\{ x' \in X : m_X(x') < m_X(x) \} < \alpha |X| \}$$

FIG. 1.2

RATIONAL CHOICE CONDITIONS AND PAIR-DOMINANT CHOICE FUNCTIONS

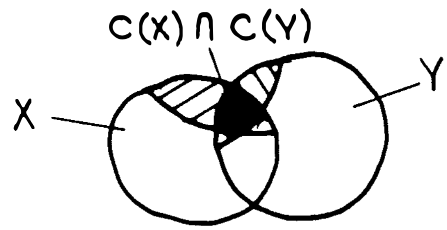
HERITANCE PROPERTY (H)

$\forall\ X' \subseteq X \subseteq \mathfrak{A} \qquad C(X') \supseteq C(X) \cap X'$



CONCORDANCE PROPERTY (C)

$\forall\ X, Y \subseteq \mathfrak{A} \qquad C(X) \cap C(Y) \subseteq C(X \cup Y)$



BINARY RELATION ($\mathcal{R}$)

$\mathcal{R} \subset \mathfrak{A} \times \mathfrak{A}$

$\mathcal{R}_x = \{y : x \mathcal{R} y\}$ - UPPER CUT OF "x" POINT

$x \mathcal{R} y \Longleftrightarrow y \in \mathcal{R}_x$ - EQUIVALENT DEFINITIONS OF BINARY RELATION

PAIR-DOMINANT CHOICE FUNCTION

$C_{\mathcal{R}}(X) \underset{df}{=} \{x : \overline{\exists}\, y \in X : y \mathcal{R} x\}$

1. SCALAR FUNCTION EXTREMIZATION

$x \mathcal{R} y \Longleftrightarrow f(x) > f(y)$

2. THRESHOLD CHOICE FUNCTION $C^{\mathcal{B}}$

$C^{\mathcal{B}}(X) = X \cap \mathcal{B}$

$\mathcal{R} = \{(x, y) \in \mathfrak{A} \times (\mathfrak{A} \setminus \mathcal{B})\}, \quad \mathcal{B} \subset \mathfrak{A}$

FIG 1.3

PROPERTIES OF PAIR-DOMINANT CHOICE FUNCTIONS

1. EVERY PAIR-DOMINANT CHOICE FUNCTION SATISFIES H CONDITION

2. EVERY CHOICE FUNCTION, SATISFIING H CONDITION CAN BE EXPANDED INTO UNION OF PAIR-DOMINANT ONES (POSSIBLE INTO INFINITE NUMBER).

$$C = \bigcup_i C_{\mathcal{R}_i}$$

<u>DEFINITION</u> C SATISFIES C_K CONDITION IFF

$$\forall X_1, \dots, X_K \in \mathfrak{X} \qquad \bigcap_{1 \le i \le K} C(X_i) \subset \bigcup_{1 \le i < j \le K} C(X_i \cup X_j)$$

3. IF C SATISFIES H AND C_K CONDITIONS IT CAN BE EXPANDED INTO UNION OF NO MORE THEN K-1 PAIR-DOMINANT CHOICE FUNCTIONS

$$C = \bigcup_{1 \le i \le K-1} C_{\mathcal{R}_i}$$

FIG 1.4

BASICAL OBJECTS

$\mathfrak{A}$ - TOTAL SET OF VARIANTS

C - CHOICE FUNCTION ON $\mathfrak{A}$

$\mathfrak{X}$ - ADMISSIBLE PRESENTATIONS. WE ASSUME ANY FINITE $X \in \mathfrak{X}$

δ - METRICS ON FINITE SUBSETS OF $\mathfrak{A}$:

$$\delta(X,Y) = \frac{\mathrm{CARD}(X \Delta Y)}{\mathrm{CARD}(X \cup Y)}$$

$X(\Theta)$ - A FAMILY OF FINITE RANDOM PRESENTATIONS $\Theta \in \mathbb{R}$ SUCH THAT $\mathbf{E}\,\mathrm{CARD}\,X(\Theta) \to \infty$, $\Theta \to \infty$

DEFINITION

CHOICE FUNCTIONS C_1 AND C_2 ARE ASYMPTOTICALLY EQUIVALENT (DEPICTED AS C_1 AEQ C_2)

$$\delta(C_1(X(\Theta)), C_2(X(\Theta))) \xrightarrow{P} 0 \quad \text{WHEN} \quad \Theta \to \infty$$

THEOREM

AEQ IS THE EQUIVALENCE RELATION ON CHOICE FUNCTIONS ON $\mathfrak{A}$

RELATION AEQ DEPENDS ON $X(\Theta)$

FIG 1.5

BASICAL CONSTRUCTIONS

$X(n)$ - SEQUENTIAL INDEPENDENT SAMPLE FROM $(\mathfrak{X}, \mathfrak{F}, \nu)$ OF VOLUME n.

ALL $\mathfrak{R} \subset \mathfrak{X} \times \mathfrak{X}$ ASSUME LAID IN $\mathfrak{F} \otimes \mathfrak{F}$. CONSEQUENTLY ALL POWERS OF SUBSETS BELOW ARE MEASURABLE.

<u>DEFINITION</u> CHOICE FUNCTIONS C_1 AND C_2 ARE ASYMPTOTICALLY INDEPENDED IF

$$\frac{\mathbf{E}\ \mathrm{CARD}\,(C_1(X) \cap C_2(X))}{\mathbf{E}\ \mathrm{CARD}\ C_i(X)} \to 0 \quad \text{WHEN } n \to \infty, \quad i = 1, 2$$

<u>DEFINITION</u> CHOICE FUNCTION C IS WELL-CONDITIONED IF

$$\frac{\mathbf{D}\ \mathrm{CARD}\ C(X)}{\mathbf{E}^2\ \mathrm{CARD}\ C(X)} \to 0 \quad \text{WHEN } n \to \infty$$

(C IS W.-C.)

<u>THEOREM</u> LET $C_1(X) \subset C_2(X)$ A.S.,

$$\frac{\mathbf{E}\ C_1(X)}{\mathbf{E}\ C_2(X)} \to 1$$ WHEN $n \to \infty$ AND ONE OF C_1 OR C_2 IS W.-C. THEN C_1 AEQ C_2.

<u>COROLLARY</u> C_2 IS W.-C. $\Rightarrow$ C_1 IS W.-C.

<u>THEOREM</u> LET C SATISFIES H PROPERTY THEN $\dfrac{\mathbf{E}\ \mathrm{CARD}\ C(X)}{n}$ NOT INCREASE WITH n.

FIG 1.6

RESULTS

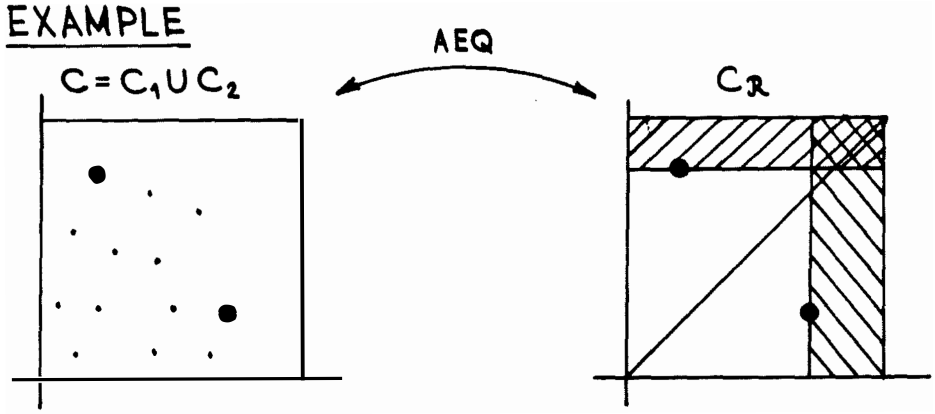
<u>THEOREM</u> LET C SATISFIES H CONDITION,

$$\lim_{n\to\infty} \frac{\mathbb{E}\,\mathrm{CARD}\, C(X)}{n} = \alpha > 0.$$

THEN THERE EXISTS $\mathfrak{B} \subset \mathfrak{A}$ SUCH THAT C AEQ $C_{\mathfrak{B}}$, $\nu(\mathfrak{B}) = \alpha$.

<u>THEOREM</u> LET $C = \bigcup_{i\in I} C_{\mathfrak{R}_i}$, $|I| < \infty$; $C_{\mathfrak{R}_i}$ ARE PAIRWISE INDEPENDENT, C IS W.-C. THEN THERE EXISTS $\mathfrak{R}$: C AEQ $C_{\mathfrak{R}}$

<u>EXAMPLE</u>



$$f : \mathfrak{A} \times \mathfrak{A} \to \mathbb{R}; \quad \mu(x) = \int_{\mathfrak{A}} f(x,y)\, d\nu(y)$$

$$\Phi(\mu) = \nu\{x : \mu(x) < \mu\}$$

<u>THEOREM</u> LET $\alpha > 0$, f IS MEASURABLE, BOUNDE Φ - CONTINUOUS.
THEN $C_{f\alpha}$ AEQ $C_{\mathfrak{B}}$, WHERE
$\mathfrak{B} = \{x : \mu(x) < \mu^*\}$, μ^* IS A ROOT OF EQUATION $\Phi(\mu) = \alpha$.

FIG 1.7

CHOICE FUNCTIONS

$(\mathfrak{A}, \mathfrak{X}, C)$

$\mathfrak{A}$ - GLOBAL SET OF ALTERNATIVES

$\mathfrak{X}$ - SET OF ADMISSIBLE PRESENTATIONS, SUBSET OF $\mathfrak{A}$

C - OPERATOR OF CHOICE

$$X \in \mathfrak{X}, \quad C(X) \subseteq X$$

CONDITION $\mathcal{H}$

$$\forall X, X' \in \mathfrak{X}, X' \subseteq X \qquad C(X) \cap X' \subseteq C(X')$$

CONDITION $\mathcal{O}$

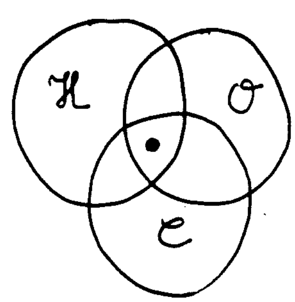
$$\forall X \in \mathfrak{X}, \forall X' \in \mathfrak{X} : C(X) \subseteq X' \subseteq X \Rightarrow C(X) = C(X')$$

CONDITION $\mathcal{C}$

$$\forall X, Y \in \mathfrak{X} \quad C(X) \cap C(Y) \subseteq C(X \cup Y)$$

$\mathcal{H} \cap \mathcal{C}$ IS THE CLASS OF NORMAL CHOICE FUNCTIONS

$$C_{\mathcal{R}}(X) = \{ x \in X \mid \neg(y \mathcal{R} x), \forall y \in X \}, \quad \mathcal{R} \subset \mathfrak{A} \times \mathfrak{A}$$



CONDITION $\mathcal{C}_K$

$$\forall X_1, \dots, X_K \in \mathfrak{X} : x \in \bigcap_{i=1}^{K} C(X_i) \Rightarrow x \in \bigcup_{1 \le i < j \le K} (C_i \cup C_j)$$

$$C \in \mathcal{H} \cap \mathcal{C}_K \Rightarrow C = \bigcup_{i=1}^{K<\infty} C_{\mathcal{R}_i}$$

Fig. 2.1

RANDOM SUBSET X

MAIN CASE : ν – PROBABILISTIC MEASURE ON $\mathfrak{X}$, $\mathfrak{G}$ – σ-ALGEBRA OF MEASURABLE SUBSETS

X – SEQUENTIAL INDEPENDENT SAMPLE OF $\mathfrak{X}$,

$$|X| = N \Leftrightarrow X = (x_1, \ldots, x_N) ,$$

$$P(x_i \in A) = \nu(A) , \quad A \in \mathfrak{G}$$

EVENTS $\{x_i \in A\}$, $\{x_j \in B\}$ ARE INDEPENDENT WHEN $i \neq j$, $A, B \in \mathfrak{G}$

N IS , GENERALLY, RANDOM .

EXAMPLES :

1. $N = n$ WITH PROBABILITY 1
2. N HAS POISSON DISTRIBUTION

Fig 2.2

MEASURABLE CHOICE FUNCTIONS AND BINARY RELATIONS

$\mathfrak{X} = \{ X \mid X$ - COUNTABLE $\}$

C - MEASURABLE $\Longleftrightarrow$ FOR EVERY SEQUENCE
$b = (b_1, b_2, \ldots)$ $\quad (b_i = 0, 1.)$
EVENT $\{ x_i \in C(X) \Longleftrightarrow b_i = 1 \}$ IS MEASURABLE

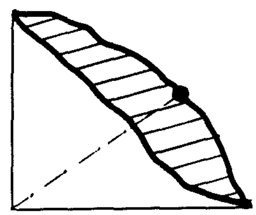
$\mathfrak{R} \subset \mathfrak{X} \times \mathfrak{X}$ IS MEASURABLE $\Longleftrightarrow$ $\mathfrak{R} \in \mathcal{Y} \otimes \mathcal{Y}$

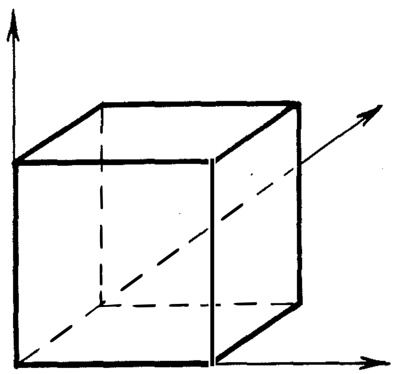
THEOREM $\mathfrak{R}$ MEASURABLE $\Longleftrightarrow$ $C_{\mathfrak{R}}$ MEASURABLE

THEOREM C_1, C_2 - MEASURABLE $\Rightarrow$ $C_1 \cup C_2$, $C_1 \cap C_2$ MEASURABLE

Fig 2.3

CLASSICAL RESULTS: INDEPENDENT CRITERIA, PARETO CONE





$$\mathfrak{X} = \mathbb{R}^m$$

$$x_i = (x_i^1, \dots, x_i^m)$$

x_i^k, x_i^j ARE INDEPENDENT WHEN $j \neq k$

$F_j(t) = P(x_1^j < t)$ – CONTINUOUS

$$ES(n) = \sum_{1 \le i_1 \le \dots \le i_{m-1} \le n} \frac{1}{i_1, i_2, \dots, i_{m-1}} \approx \frac{\ln^{m-1} n}{(m-1)!}$$

$$DS(n) \sim ES(n)$$

$m = 2$: $S(n)$ IS ASIMPTOTICALLY NORMAL

Fig 2.4

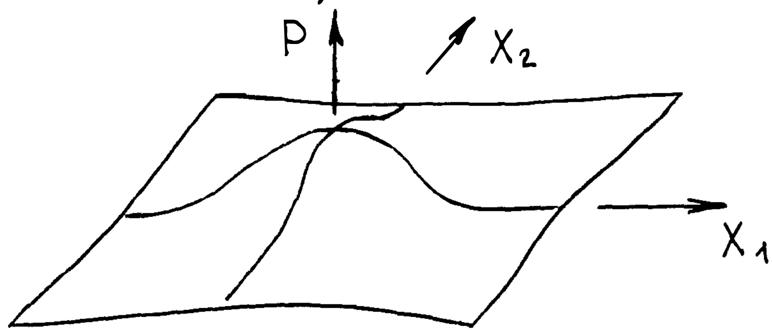
ANOTHER CLASSICAL RESULTS: PARETO CONE

1. MULIDIMENSIONAL NORMAL DISTRIBUTION

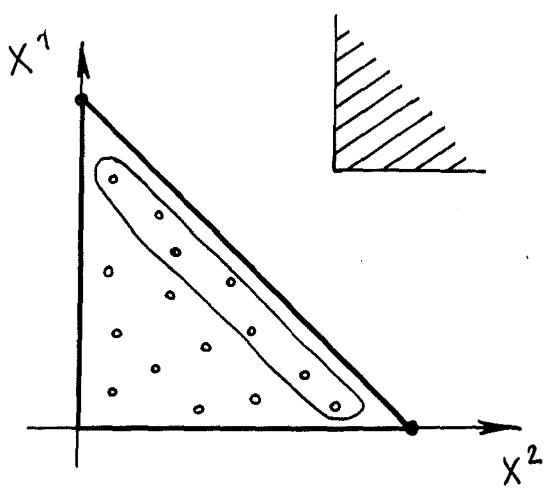
$$p(\bar{x}) = c\, e^{-\frac{1}{2}(x, Ax)}$$

A – $m \times m$ MATRIX, POSITIVELY DEFINED

$$E\, S(n) \sim c \ln^{m-1} n$$



2. UNIFORM DISTRIBUTION IN SIMPLEX $X^i \geqslant 0,\ \sum X^i \leqslant 1$



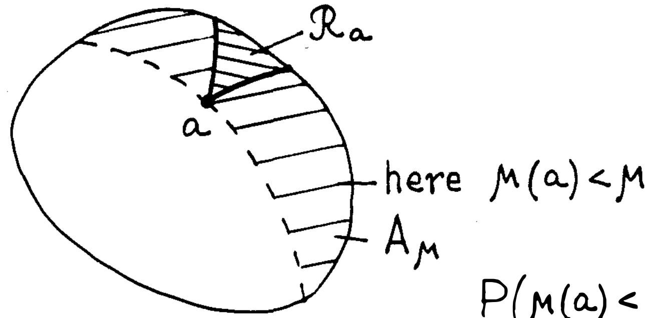
$$E\, S(n) \sim n^{\frac{1}{2}}, \quad m = 2$$

$$E\, S(n) \sim n^{\frac{m-1}{m}}$$ – GENERAL CASE

Fig 2.5

MEASURE OF UPPER CUT

$$\mu(a) = P(x \mathcal{R} a) = \nu(\mathcal{R}_a)$$



$$P(\mu(a) < \mu) = B(\mu) = \nu(A_\mu)$$

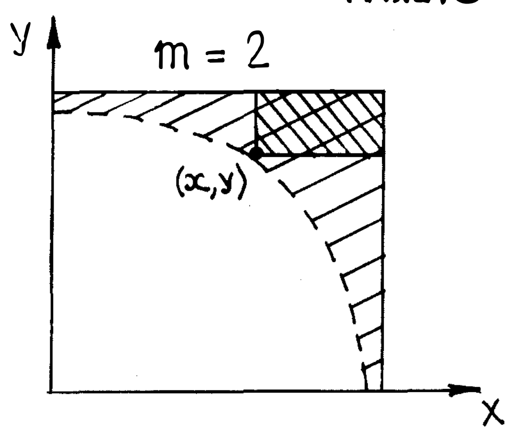
$$ES(n) = n \int_0^1 (1-\mu)^{n-1} dB(\mu)$$

$B(\mu) \longmapsto e(n) = ES(n)$: MELLIN TRANSFORMATION

PROPERTIES OF MELLIN TRANSFORMATION

1. $B(\mu) = B_1(\mu) + B_2(\mu)$
   $B_2 = o(B_1) \Rightarrow e(n) = e_1(n) + e_2(n)$, $e_2 = o(e_1)$
2. MELLIN'S TRANSFORMATION $\sim$ LAPLACE TRANSFORMATION
3. $B(\mu) = \mu^\alpha h(\mu)$, WHERE $\lim\limits_{\mu \to 0} \frac{h(t\mu)}{h(\mu)} = 1$ FOR ANY $t \Rightarrow$
   $\Rightarrow e(n) \sim n^{1-\alpha} h\left(\frac{1}{n}\right)$.

EXAMPLE: CLASSICAL CASE -- m INDEPENDENT CRITERIA, PARETO CONE



$$\mu((x,y)) = (1-x)(1-y)$$

$$A_\mu = \{(x,y) : (1-x)(1-y) < \mu\}$$

$$B_\mu = \mu - \mu \ln \mu \Rightarrow e(n) \approx \ln n$$

Fig 2.6

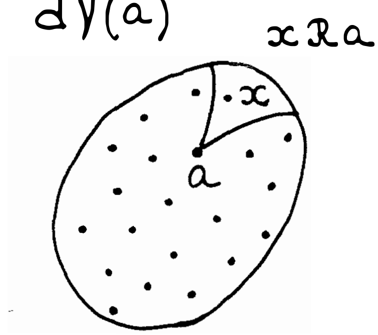
FORMULAS FOR $ES(N)$ FOR BINARY RELATION $\mathcal{R}$ ON $\mathfrak{X}$

X – SEQUENTIAL INDEPENDENT N-VOLUME SAMPLE.

$$S(N) = \sum_{i=1}^{N} \chi_i(N)$$

$$ES(N) = EN\chi_1(N) = \sum_{n=0}^{\infty} P(N=n)\cdot n P(\chi_1(n)=1)$$

$$P(\chi_1(n)=1) = \int_{\mathfrak{X}} (1-P(x\mathcal{R}a))^{n-1} d\nu(a)$$



$$\Phi(z) = \sum_{n=1}^{\infty} P(N=n) z^n$$

$$ES(N) = \int_{\mathfrak{X}} \frac{d\Phi}{dZ}(1-P(x\mathcal{R}a))d\nu(a)$$

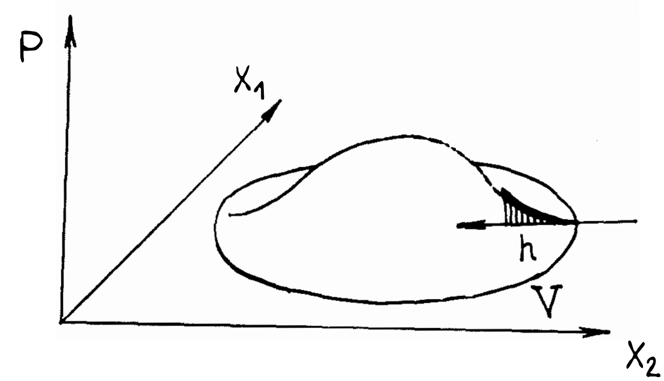
$$\mu(a) = P(x\mathcal{R}a) = \nu(\mathcal{R}_a)$$

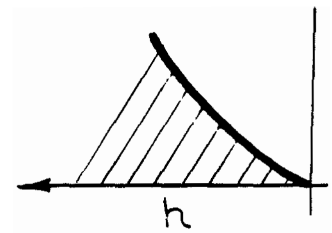
$$ES(N) = \int_{\mathfrak{X}} \frac{d\Phi}{dZ}(1-\mu(a))d\nu(a) =$$

$$= \int_0^1 \frac{d\Phi}{dZ}(1-\mu)dP(\mu(a)<\mu)$$

Fig. 2.7

CASE OF COMPACT SUPPORT IN GENERAL POSITION AND RELATION DEFINED BY SHARP CONE K

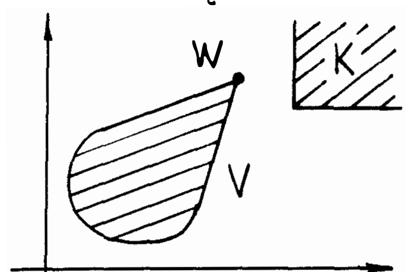


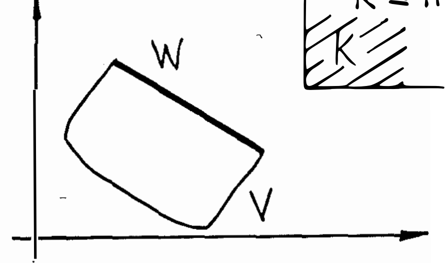


MEASURE IS DEFINED BY DENSITY $p(x)$, WITH SUPPORT V - COMPACT BODY WITH PIECEWISE SMOOTH ~~BOARD~~ Boundary AND $p(x) = b(x) \cdot h^s(x)$, WHERE $b(x) > 0$ ON V, $s > -1$, $h(x)$ - DISTANCE FROM x TO ∂V

$$W = \{ x \mid \mu(x) = 0 \} \quad \text{- WEAK OPTIMA}$$

$$W = \bigcup W_i^k \quad \text{- UNION ON STRATA WITH DIMENSION } k \leq m-1$$





IN GENERAL CASE THERE EXISTS STRATUM W_j^k, SUCH THAT THERE EXISTS $w \in W_j^k$: $T_w W_j^k \cap (K + w) = w$

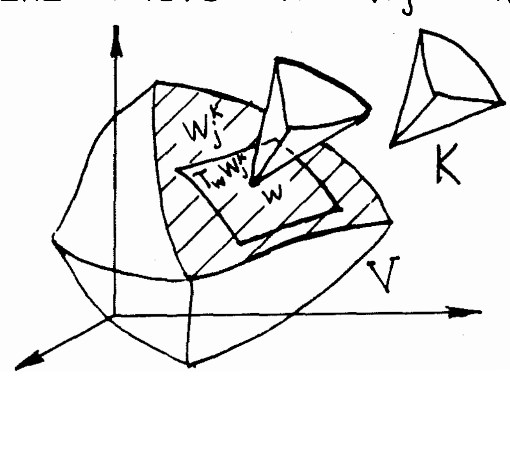
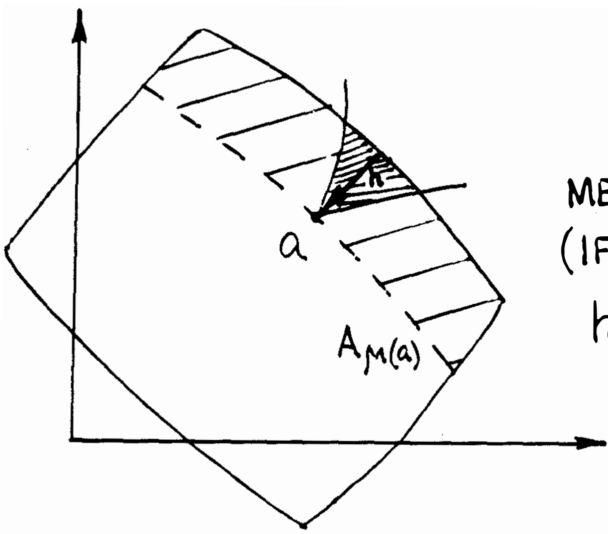


Fig 2.8

LET K BE MAXIMAL OF DIMENSIONS OF SUCH STRATA. THEN $ES(n) \sim n^{\frac{K}{m+s}}$



$\mu(a) \sim h^{m+s}$

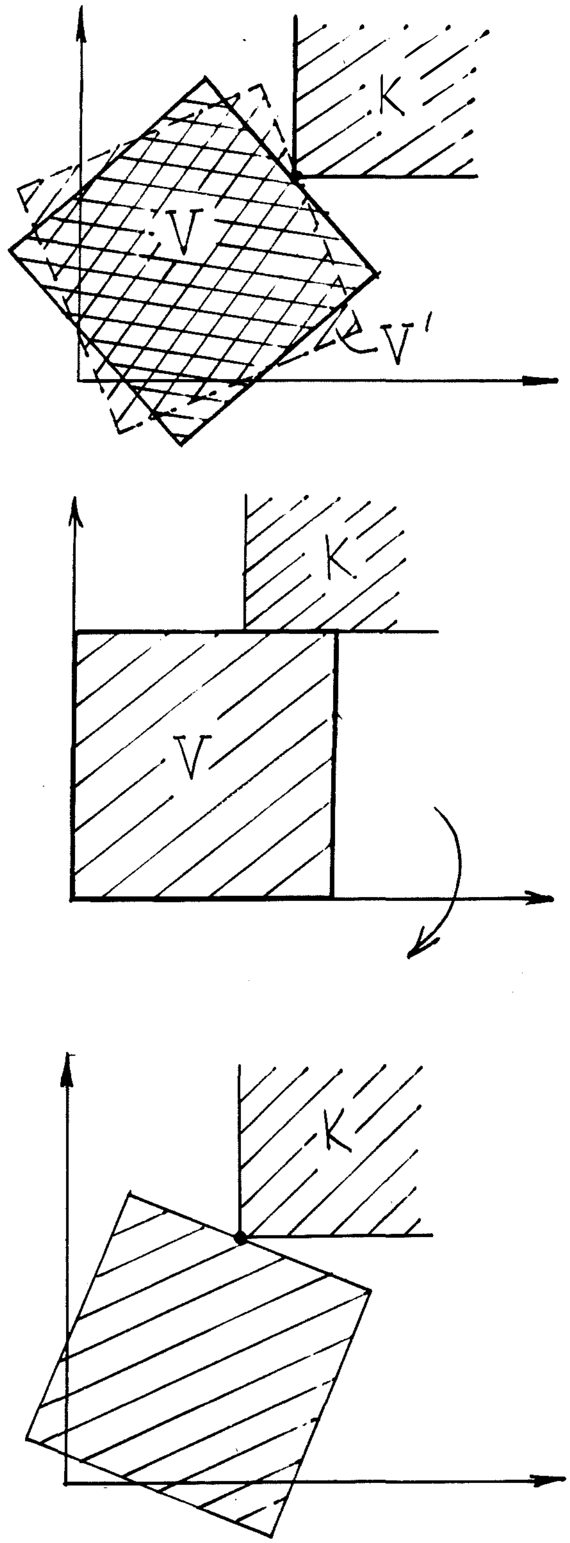
MEASURE OF $A_{\mu(a)} \sim h^{1+s}$ (IF $m-k=1$, GENERALLY - h^{m-k+s})

$$h \sim \mu^{\frac{1}{m+s}}; \qquad B(\mu) \sim h^{1+s} \sim \mu^{\frac{1+s}{m+s}}$$

$$n \int (1-\mu)^{n-1} dB(\mu) \sim n \cdot n^{-\frac{1+s}{m+s}} = n^{\frac{m-1}{m+s}}$$

Fig 2.8 (continuation)

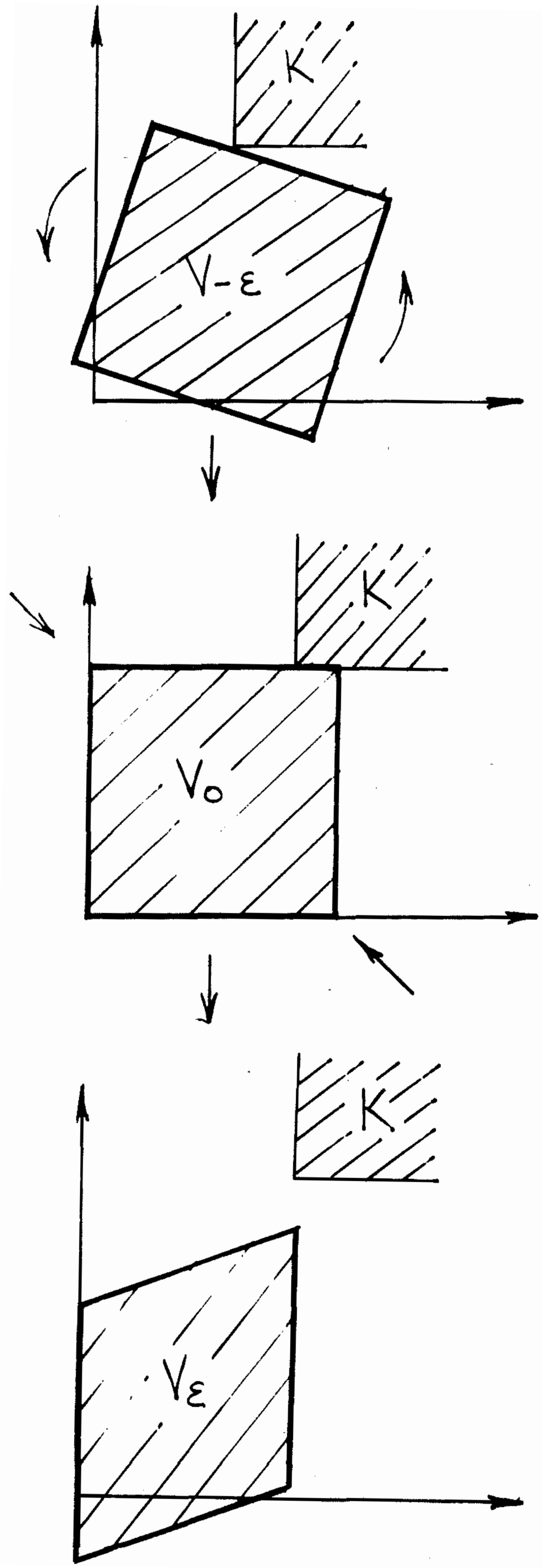
GENERAL CASE: STABILITY AND GENERICITY



STABILITY

Fig 2.9

ASYMPTOTICAL SERIES FOR $e(n)$



$e(n) \sim n^{\frac{1}{2}}$

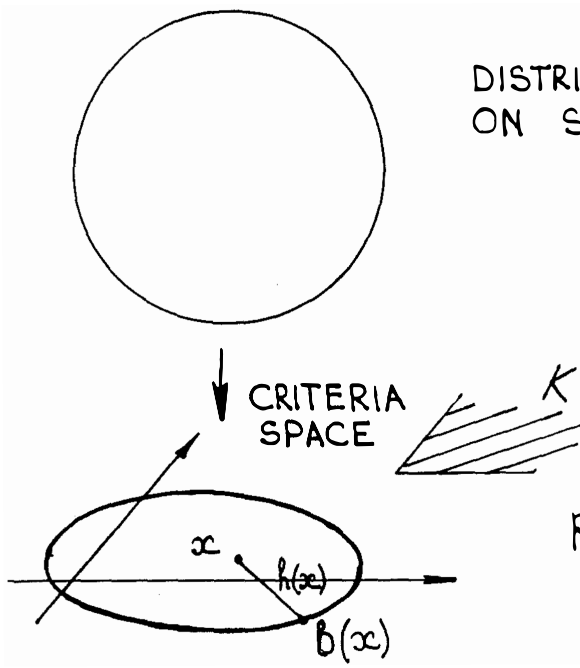
$e(n) \sim \ln n$

$e(n) \sim n^0$

Fig 2.10

EXAMPLES

1.



DISTRIBUTION IS UNIFORM ON SPHERE

$$p(x) = b(x) \cdot h^{-\frac{1}{2}}(x)$$

$\Rightarrow$ $S(n)$ IS ASYMPTOTICALLY NORMAL,

$$ES(n) \sim DS(n) \sim n^{\frac{2-1}{2-\frac{1}{2}}} = n^{\frac{2}{3}}$$

2.

$$p(x) = b(x) \cdot h(x)$$

$S(n)$ IS ASYMPTOTICALLY NORMAL,

$$ES(n) \sim DS(n) \sim n^{\frac{2-1}{2+1}} = n^{\frac{1}{3}}$$

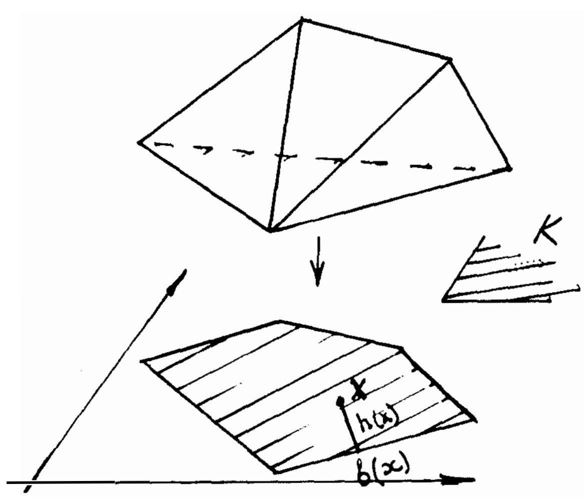
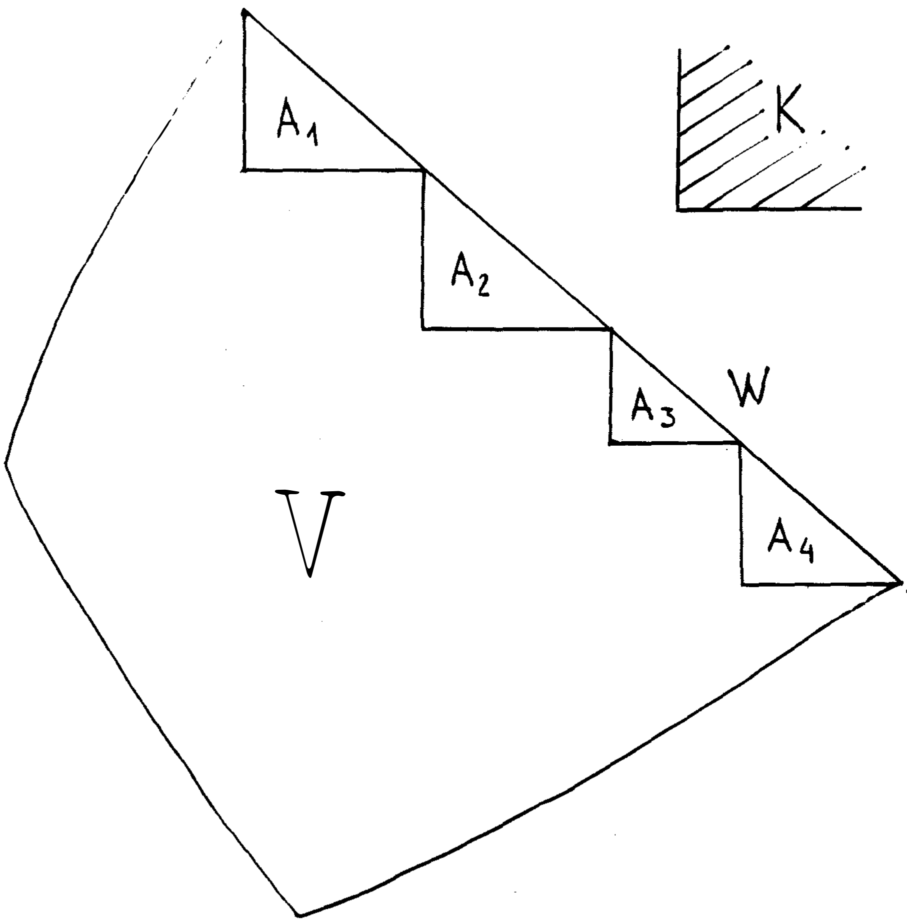


Fig 2.11

DISTRIBUTION : GENERAL CASE



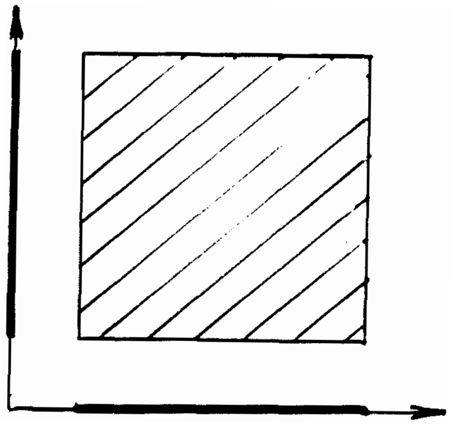
IDEA OF PROOF

$S(n) \approx \sum_i \text{CARD}\,(C(X) \cap A_i)$

CARD $(C(X) \cap A_i)$ ARE INDEPENDENT.
WE CAN ESTIMATE FOURTH MOMENTS OF CARD $(C(X) \cap A_i)$ AND BY LINDEBERG'S THEOREM SHOW, THAT $S(n)$ IS ASYMPTOTICALLY NORMAL WITH $ES(n) \sim DS(n) \sim n^{\frac{K}{m+s}}$

Fig 2.12

SPECIAL CASE: DIRECT PRODUCTS



$$\mathfrak{X} = \mathfrak{X}' \times \mathfrak{X}''$$

$$P = P' \times P''$$

$$\mathcal{R} = \mathcal{R}' \times \mathcal{R}'' \iff$$

$$(x', x'')\, \mathcal{R}\, (y', y'') \iff$$

$$x' \mathcal{R}' y' \, , \; x'' \mathcal{R}'' y''$$

$$\mu(x', x'') = \mu'(x')\; \mu''(x'')$$

$$\Rightarrow \; E\,S(n) = n \int (1 - \mu(X)\, dP(X) =$$

$$= n \iint (1 - \mu' \mu'')\, dP'(\mu'(x') < \mu')\, dP''(\mu''(x'') < \mu'') \, .$$

$$(1 - \mu' \mu'') = ((1 - \mu') + (1 - \mu'') - (1 - \mu')(1 - \mu''))$$

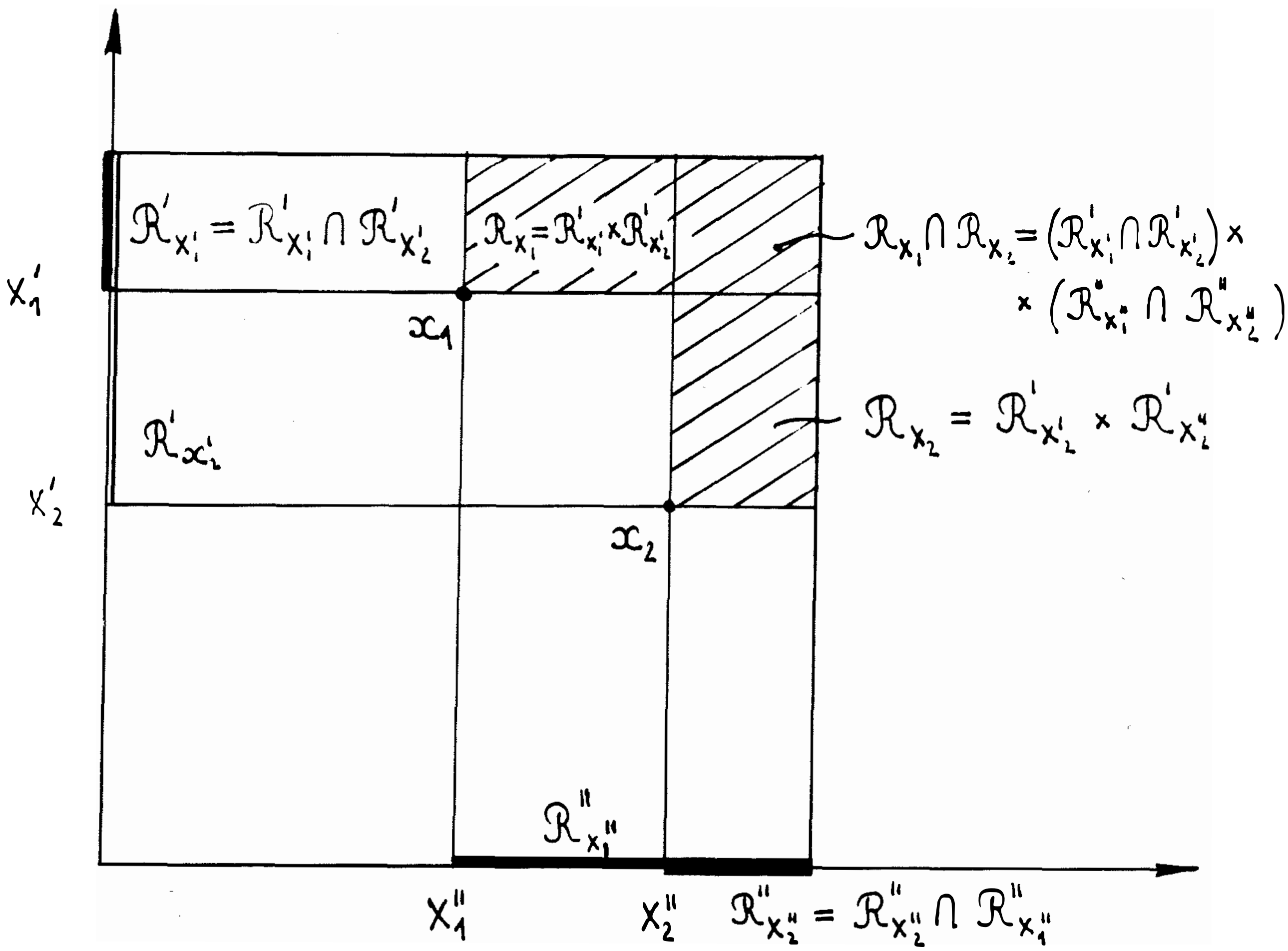
THEOREM THE KNOWLEDGE OF $\mathbf{e}'(n)$, $\mathbf{e}''(n)$ YIELD KNOWLEDGE OF $\mathbf{e}(n)$

IF $\mathfrak{X}'' = [0,1]$, P'' IS UNIFORM ON $[0,1]$, $\mathcal{R}'' = \{<\}$, THEN

$$e(n) = \sum_{i=1}^{n} \frac{e'(i)}{i}$$

Fig 2.13

EXAMPLE: $\mathbb{R}^2$, INDEPENDENT CRITERIA, PARETO CONE



$$(1-\mu) = 1 - \mu'\mu''$$

$$1 - \text{MEASURE}\left(\mathcal{R}_{x_1} \cup \mathcal{R}_{x_2}\right) = 1 - \text{MEASURE}\left(\mathcal{R}_{x_1}\right) - \text{MEASURE}\left(\mathcal{R}_{x_2}\right) + $$
$$+ \text{MEASURE}\left(\mathcal{R}_{x_1} \cap \mathcal{R}_{x_2}\right)$$

Fig 2.14

SPECIAL CASE: INFINITE PRESENTATION

$\mathfrak{A}$ - COUNTABLE (INFINITE)

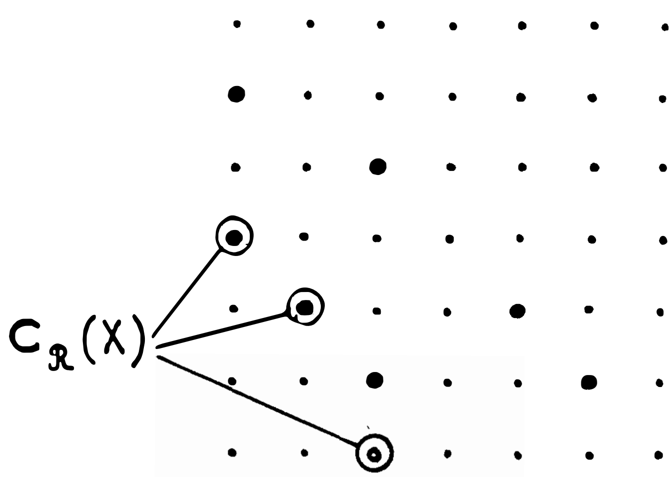
$X \subset \mathfrak{A}$ - RANDOM

EVENTS $\{a_1 \in X\}$, $\{a_2 \in X\}$ ARE INDEPENDED,

WITH $\mathbb{P}(a \in X) = p(a)$

IF $\sum_{a \in \mathfrak{A}} p(a) = \infty$, THEN card $X = \infty$ A.S.

EXAMPLE



IF $p(a) \equiv p$, THEN $S < \infty$ A.S. AND

$$\mathbf{E} S \sim \ln(1/p)$$

Fig 2.15

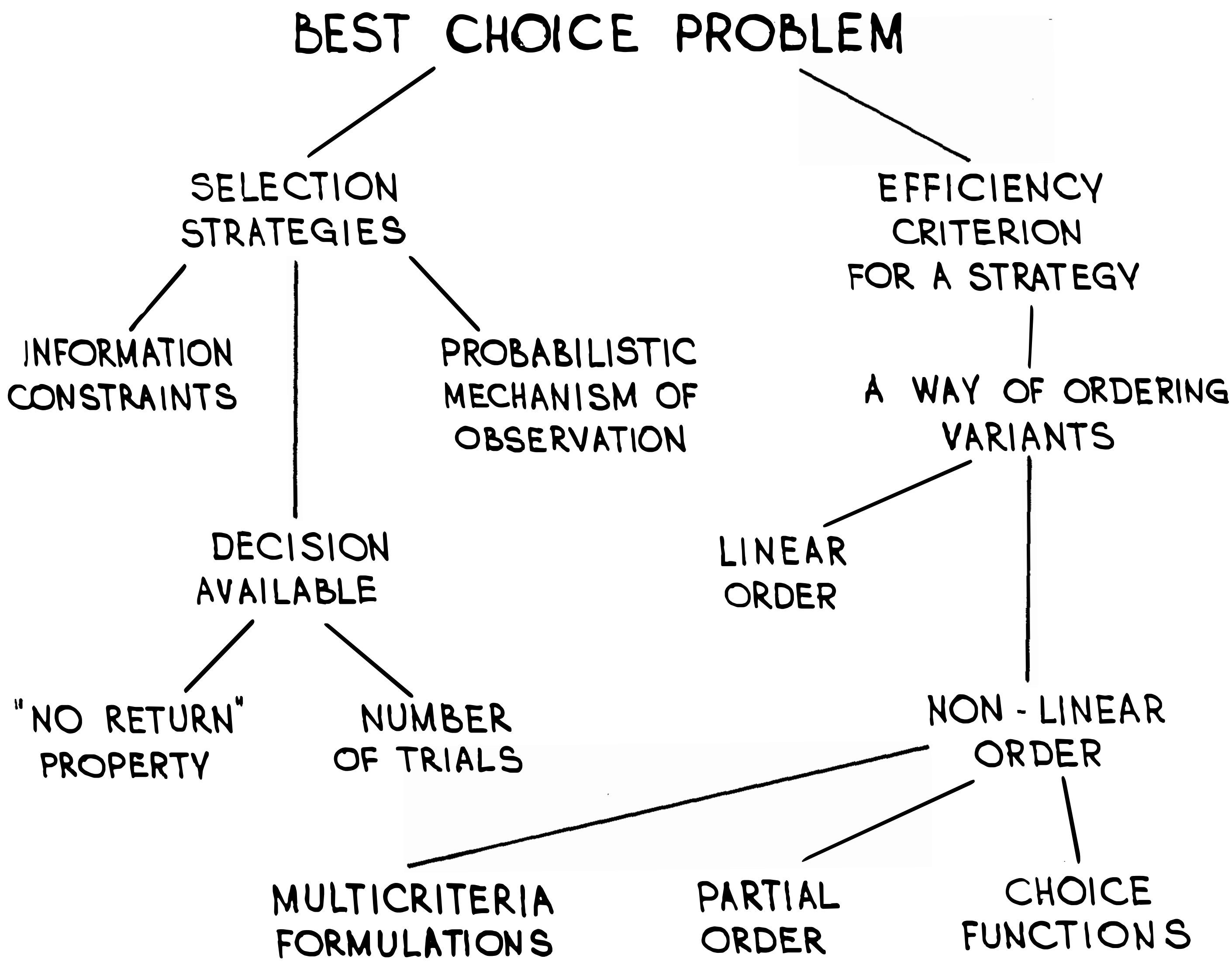

BEST CHOICE PROBLEM
SELECTION STRATEGIES
EFFICIENCY CRITERION FOR A STRATEGY
INFORMATION CONSTRAINTS
PROBABILISTIC MECHANISM OF OBSERVATION
A WAY OF ORDERING VARIANTS
DECISION AVAILABLE
LINEAR ORDER
"NO RETURN" PROPERTY
NUMBER OF TRIALS
NON - LINEAR ORDER
MULTICRITERIA FORMULATIONS
PARTIAL ORDER
CHOICE FUNCTIONS

FIG. 3.1

THE CLASSICAL PROBLEM

$X_1, \dots, X_N$ – ABSOLUTE RANKS, A RANDOM PERMUTATION OF $1, \dots, N$

$Y_1, \dots, Y_N$ – RELATIVE RANKS

t – STOPPING RULE BASED ON Y'S

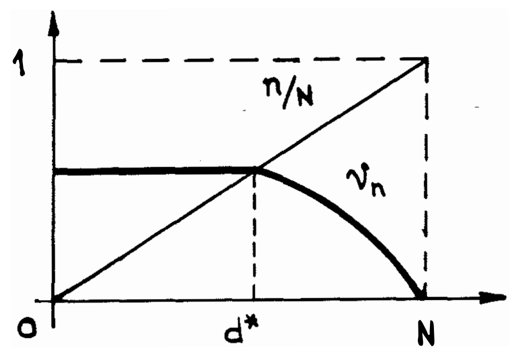
THE PROBLEM:

$$\mathcal{E}(t) \underset{def}{=} P\{X_t = 1\} \xrightarrow[t]{} \text{MAX}$$

THE Y'S ARE INDEPENDENT AND $P\{Y_n = k\} = n^{-1}$, $k \le n \le N$

$$P\{X_N = 1 \mid Y_N = 1\} = \frac{n}{N}$$

ν_n – THE OPTIMAL CONTINUATION VALUE



THE OPTIMAL RULE

$$t^* = \min\{n \geq d^* \mid Y_n = 1\}, \quad (\min \emptyset = N)$$

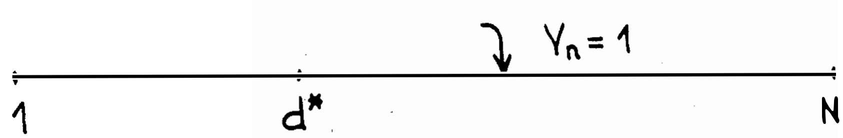


FIG 3.2

A THRESHOLD RULE

$$t_d \underset{\text{def}}{=} \text{MIN} \{ n \mid n \geqslant d \, , Y_n = 1 \} \, ,$$

$$\mathcal{E}(t_d) = \frac{d-1}{N} \sum_{K=d}^{N} \frac{1}{K-1}$$

THE OPTIMAL d^* SATISFIES

$$\sum_{K=d^*+1}^{N} \frac{1}{K-1} \leqslant 1 < \sum_{K=d^*}^{N} \frac{1}{K-1}$$

ASYMPTOTICS :

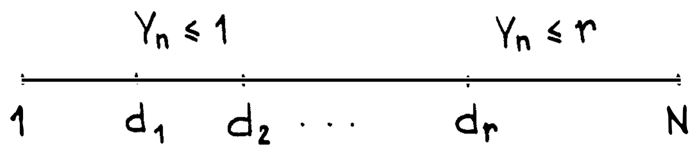
$$\lim_{N \to \infty} \frac{d^*}{N} = \lim_{N \to \infty} \mathcal{E}(t^*) = e^{-1} = 0{,}368\ldots$$

FIG 3.3

THE GUSEIN - ZADE'S PROBLEM

$$\mathcal{E}(t) = P\{X_t \leq r\} \to \text{MAX}$$

OPTIMAL STOPPING RULE :



A THRESHOLD RULE

$$t_d = \text{MIN}\{n \mid n \geq d,\ Y_n \leq r\}$$

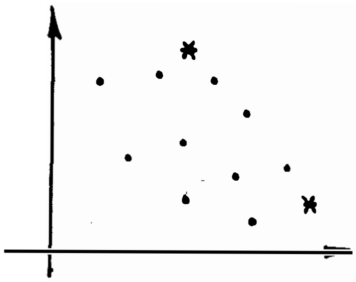
$$\mathcal{E}(t_d) = \frac{r}{N} \binom{d-1}{r} \sum_{n=d}^{N} \binom{n-1}{r}^{-1}$$

ASYMPTOTICS :

$$\lim_{N\to\infty} \mathcal{E}(t_{d^*}) = \lim_{N\to\infty} \frac{d^*}{N} = \left(\frac{1}{r}\right)^{1/(r-1)}$$

FIG 3.4

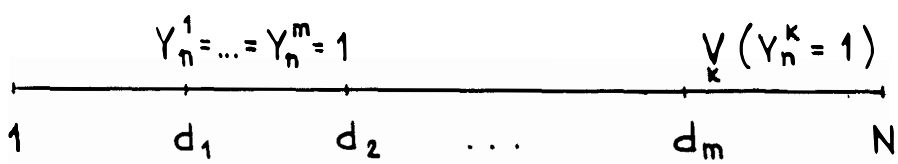
COLLECTIVE - EXTREMAL CHOICE



$$\mathbf{X}_n = \begin{pmatrix} X_n^1 \\ \vdots \\ X_n^m \end{pmatrix} \qquad \mathbf{Y}_n = \begin{pmatrix} Y_n^1 \\ \vdots \\ Y_n^m \end{pmatrix}$$

$$\mathcal{E}(t) = \mathcal{P}\left\{ \bigvee_{K=1}^{m} \left(X_t^K = 1 \right) \right\}$$

THE OPTIMAL STOPPING RULE :



A THRESHOLD RULE :

$$t_d = \text{MIN}\left\{ n \mid n \geq d, \bigvee_{K=1}^{m} \left(Y_n^k = 1 \right) \right\}.$$

ASYMPTOTICS :

$$\lim_{N\to\infty} \mathcal{E}(t^*) = \lim_{N\to\infty} \mathcal{E}(t_{d^*}) = \lim_{N\to\infty} \frac{d^*}{N} = \left(\frac{1}{m}\right)^{1/(m-1)}$$

FIG 3.5

CHOICE FUNCTION

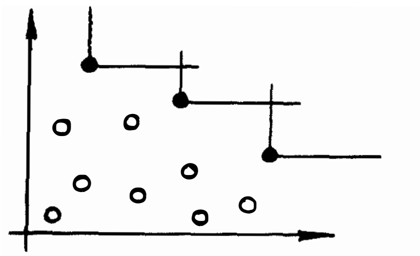
$(\mathfrak{A}, \mathfrak{X}, C)$

$\mathfrak{A}$ – THE GLOBAL SET OF ALTERNATIVES

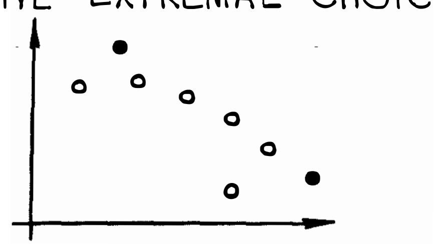
$\mathfrak{X}$ – A SYSTEM OF ADMISSIBLE SUBSETS $\mathfrak{A}$, CHOICE CONTEXTS

C – THE CHOICE OPERATOR: $C(X) \subset X$ FOR $X \in \mathfrak{X}$

PARETO OPTIMALITY



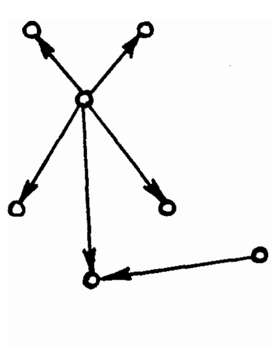
COLLECTIVE - EXTREMAL CHOICE



GUSEIN - ZADE'S CHOICE FUNCTION

$\mathfrak{A} = \{1, \dots, N\}$, $C(a_1, \dots, a_n) = \{a_k \mid \text{RANK } a_k \text{ AMONG } a_1, \dots, a_k \leq r\}$, $r > 1$

PARTIAL ORDER ON A FINITE SET



$C(a_1, \dots, a_n)$ NONDOMINATED VERTICES IN SUBGRAPH $(a_1, \dots, a_n) \subset \mathfrak{A}$

FIG 3.6

GENERAL STATEMENT

$\mathfrak{X}$ - A SAMPLING SPACE

$(x_1, \dots, x_N)$ - A RANDOM SAMPLE

C - CHOICE FUNCTION

t - A STOPPING RULE BASED ON x'S

$$\mathcal{E}(t) = P\{x_t \in C(A_1, \dots, A_N)\} \to \text{MAX}$$

A THRESHOLD STOPPING RULE:

$$t_d = \text{MIN}\{n \mid n \geq d,\ x_n \in C(x_1, \dots, x_n)\}$$

CONDITION H

$$x_1 \in C(x_1, \dots, x_n) \Rightarrow x_1 \in C(x_1, \dots, x_{n-1})$$

CONDITION O

$$x_1 \notin C(x_1, \dots, x_n) \Rightarrow C(x_2, \dots, x_n) = C(x_1, \dots, x_n)$$

$$S_n = \text{CARD } C(x_1, \dots, x_n)$$

THEOREM 1. IF $C \in H$, $S_n = r$ FOR ALL $r \leq n \leq N$ THEN

$$\mathcal{E}(t_d) = \frac{r}{N}\binom{d-1}{r}\sum_{n=d}^{N}\binom{n-1}{r}^{-1}, \quad d > r.$$

THEOREM 2. IF $C \in H \cap O$, $S_n \leq m$ FOR ALL $n = 1, \dots, N$, $E S_n = r$,

$$\text{MAX } \mathcal{E}(t_d) \geq \begin{cases} \dfrac{r}{m}\left(\dfrac{1}{m}\right)^{1/(m-1)}, & \text{IF } m > 1 \\ r e^{-1}, & \text{IF } m = 1 \end{cases}$$

FIG 3.7

PARETO OPTIMALITY

$$\xi_d = \text{CARD}\{ n \mid d \leq n \leq N ; \; x_n \in C(x_1, \dots, x_N) \}$$

$$\eta_d = \text{CARD}\{ n \mid d \leq n \leq N , \; x_n \in C(x_1, \dots, x_{d-1}, x_N) \}$$

THEOREM
$$\mathcal{E}(t_d) = E \, \frac{\xi_d}{\eta_d}$$

ASYMPTOTICS :

$$\lim_{N \to \infty} \; \max_{d} \; \mathcal{E}(t_d) = 1$$

FIG 3.8